DOCUMENT RESUME

ED 321 317 CS 507 208

AUTHOR	Studdert-Kennedy, Michael, Ed.
TITLE	Status Report on Speech Research, January-June 1989.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, Md.; National Inst. of Health (DHHS), Bethesda, MD. Biomedical Research Support Grant Program.; National Inst. of Neurological and Communicative Disorders and Stroke (NIH), Bethesda, Md.; National Science Foundation, Washington, D.C.
REPORT NO	SR-97/98
PUB DATE	89
CONTRACT	NO1-HD-5-2910
GRANT	BNS-8520709; HD-01994; NS07237; NS13617; NS13870; NS18010; NS24655; NS25455; RR-05596
NOTE	265p.; For the July-December 1989 status report, see CS 507 209.
PUB TYPE	Collected Works - General (020) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC11 Plus Postage.
DESCRIPTORS	*Auditory Perception; Communication Research; Dyslexia; *Language Processing; Language Research; Music Theory; *Phonetics; Reading Comprehension; Reading Difficulties; *Speech; *Speech Communication
IDENTIFIERS	*Speech Perception; Speech Research; Vocalization

ABSTRACT

One of a series of semiannual reports, this publication contains 14 articles which report the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. The titles of the articles and their authors are as follows: "Gestural Structure and Phonological Patterns" (Catherine P. Browman and Louis Goldstein); "The Cricothyroid Muscle in Voicing Control" (Anders Lofqvist and others); "Growth of Surface Waves with Application to the Mucosal Wave of the Vocal Folds" (R. S. McGowan); "Effects of Preceding Context on Discrimination of Voice Onset Times" (Bruno H. Repp and Hwei-Bing Lin); "Can Speech Perception be Influenced by Simultaneous Presentation of Print?" (Ram Frost and others); "Detectability of Words and Nonwords in Two Kinds of Noise" (Bruno H. Repp and Ram Frost); "Discovering Phonetic Coherence in Acoustic Patterns" (Catherine T. Best and others); "Reception of Language in Broca's Aphasia" (Donald Shankweiler and others); "Syntactic Comprehension in Young Poor Readers" (Suzanne T. Smith and others); "Phonological Deficiencies: Effective Predictors of Future Reading Problems" (Virginia A. Mann and Patricia Ditunno); "Consideration of the Principles of the International Phonetic Alphabet" (Arthur S. Abramson); "Language Specificity in Lexical Organization: Evidence from Deaf Signers' Lexical Organization of ASL and English" (Vicki L. Hanson and Laurie B. Feldman); "Anomalous Bimanual Coordination among Dyslexic Boys" (Marshall Gladstone and others); "Expressive Microstructure in Music: A Preliminary Perceptual Assessment of Four Composers' Pulses'" (Bruno H. Repp). An appendix lists DTIC and ERIC numbers for publications in this series since 1970. (SR)

ED321317

Haskins Laboratories Status Report on
Speech Research

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
A. DADOURIAN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

CS507208

SR-97/98
JANUARY-JUNE 1989

BEST COPY AVAILABLE

Status Report on

Speech Research

SR-97/98
January - June 1989

Haskins Laboratories
New Haven, Connecticut 06511

Distribution Statement

Editor
Michael Studdert-Kennedy

Status Report Publisher
Yvonne Manning

Technical Illustrator
Zefang Wang

This publication reports the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications.

Distribution of this document is unlimited. The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.

Correspondence concerning this report should be addressed to the Editor at the address below:

Haskins Laboratories
270 Crown Street
New Haven, Connecticut 06511-6695

Acknowledgment

The research reported here was made possible in part
by support from the following sources:

National Institute of Child Health and Human Development:

Grant HD-01994
Contract NO1-HD-5-2910

National Institutes of Health:

Biomedical Research Support Grant RR-05596

National Science Foundation

Grant BNS-8520709

National Institute of Neurological and Communicative Disorders and Stroke

Grant NS 13870
Grant NS 13617
Grant NS 18010
Grant NS 24655
Grant NS 07237
Grant NS 25455

Investigators

Arthur Abramson*
Peter J. Alfonso*
Thomas Baer
Eric Bateson*
Fredericka Bell-Berti*
Catherine T. Best*
Susan Brady*
Catherine P. Browman
Franklin S. Cooper*
Stephen Crain*
Robert Crowder*
Lois Dreyer*
Alice Faber†
Laurie B. Feldman*
Janet Fodor*
Anne Fowler
Carol A. Fowler*
Louis Goldstein*
Carol Gracco†
Vincent Gracco
Vicki L. Hanson*
Katherine S. Harris*
Leonard Katz*
Rena Arens Krakow*
Andrea G. Levitt*
Alvin M. Liberman*
Isabelle Y. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist*
Jochen Machetanz[1]
Virginia H. Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Patrick W. Nye
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Michael Studdert-Kennedy*
Koichi Tsunoda[2]
Michael T. Turvey*
Douglas Whalen

Technical/Administrative Staff

Philip Chagnon
Alice Dadourian
Michael D'Angelo
Betty J. DeLise
Vincent Gulisano
Donald Hailey
Raymond C. Huey*
Marion MacEachron*
Yvonne Manning
Joan Martinez
Lisa Mazur
Vance Maverick
William P. Scully
Richard S. Sharkany
Edward R. Wiley
Zefang Wang

Students*

Jennifer Aydelott
Dragana Barac
Paola Bellabarba
Melanie Campbell
Sandra Chiang
André Cooper
Margaret Hall Dunn
Terri Erwin
Elizabeth Goodell
Joseph Kalinowski
Deborah Kuglitsch
Herbert Leass
Simon Levy
Katrina Lukatela
Diana Matson
Gerald W. McRoberts
Pamela Mermelstein
Salvatore Miranda
Maria Mody
Weijia Ni
Mira Peter
Nian-qi Ren
Christine Romano
Arlyne Russo
Richard C. Schmidt
Jeffrey Shaw
Caroline Smith
Jennifer Snyder
Robin Seider Story
Rosalind Thornton
Mark Tiede
Qi Wang
Yi Xu

*Part-time
†NRSA Training Fellow
[1] Visiting from Georg-August-Universitat Gottingen, Germany
[2] Visiting from University of Tokyo, Japan

Contents

Gestural Structures and Phonological Patterns
Catherine P. Browman and Louis Goldstein 1

The Cricothyroid Muscle in Voicing Control
Anders Löfqvist, Thomas Baer, Nancy S. McGarr, and Robin Seider Story.............. 25

Growth of Surface Waves with Application to the Mucosal Wave of the Vocal Folds
R. S. McGowan 41

Effects of Preceding Context on Discrimination of Voice Onset Times
Bruno H. Repp and Hwei-Bing Lin 51

Can Speech Perception be Influenced by Simultaneous Presentation of Print?
Ram Frost, Bruno H. Repp, and Leonard Katz 69

Detectability of Words and Nonwords in Two Kinds of Noise
Bruno H. Repp and Ram Frost 87

Discovering Phonetic Coherence in Acoustic Patterns
Catherine T. Best, Michael Studdert-Kennedy, Sharon Manuel, and Judith Rubin-Spitz 95

Reception of Language in Broca's Aphasia
Donald Shankweiler, Stephen Crain, Paul Gorrell, and Betty Tuller 121

Syntactic Comprehension in Young Poor Readers
Suzanne T. Smith, Paul Macaruso, Donald Shankweiler, and Stephen Crain 147

Phonological Deficiencies: Effective Predictors of Future Reading Problems
Virginia A. Mann and Patricia Ditunno 171

Consideration of the Principles of the International Phonetic Alphabet
Arthur S. Abramson 191

Language Specificity in Lexical Organization: Evidence from Deaf Signers' Lexical Organization of ASL and English
Vicki L. Hanson and Laurie B. Feldman 195

Anomalous Bimanual Coordination Among Dyslexic Boys
Marshall Gladstone, Catherine T. Best, and Richard J. Davidson 213

Expressive Microstructure in Music: A Preliminary Perceptual Assessment of Four Composers' "Pulses"
Bruno H. Repp 235

Appendix 261

Status Report on

Speech Research

SR-97/98 January - June, 1989

Gestural Structures and Phonological Patterns*

Catherine P. Browman and Louis Goldstein[†]

INTRODUCTION

In this paper, we present a particular view of phonology, one in which lexical items are construed as characterizing the activity of the vocal tract (and its articulators) as a sound-producing system. Central to this characterization is the concept of dynamically defined articulatory gestures. Such gestures are, we argue, useful primitives for characterizing phonological patterns as well as for analyzing the activity of the vocal tract articulators. Gestural structures can function as lexical representations because distinctiveness is captured by these structures and phonological processes can be seen as operating on them. In addition, the gestural approach has two benefits: it serves to *analyze* phonological patterns according to their source (articulatory, acoustic, perceptual, etc.), and it captures the underlying *unity* among apparently disparate patterns, not only within phonology but also among phonology, perception, and production. These benefits lead to a simpler, more general account of a number of phenomena.

A gestural lexicon is part of the computational model of speech production that we are currently developing with Elliot Saltzman, Philip Rubin and others at Haskins Laboratories (Browman, Goldstein, Kelso, Rubin, & Saltzman, 1984; Browman, Goldstein, Saltzman, & Smith, 1986; Saltzman, Rubin, Goldstein, & Browman, 1987). Our notion of gesture is based on the concept of coordinative structures (Fowler, Rubin, Remez, & Turvey, 1980) as developed in the task dynamic model (Saltzman, 1986; Saltzman & Kelso, 1987), and is consistent with the view of Liberman and Mattingly (1985) that "gestures...have characteristic invariant properties...as the more remote structures that control the [peripheral] movements" (p. 23). Briefly, the basic assumptions are (1) that a primary task in speaking is to control the coordinated movement of groups of articulators (rather than the individual movements of individual articulators), and (2) that these coordinated movements can be characterized using dynamical equations. Each gesture has its own characteristic coordinative pattern.

The hypothesis that the lexicon is composed of dynamically specified gestures has several implications for the motor theory and modularity claims. First, since the same lexicon is assumed to be accessed when an individual is speaking or listening, the hypothesis implies that a listener ultimately recovers the set of gestures that are part of a given lexical entry. However, the motor theory is a specific proposal about the *mechanism* of recovery (among several possible proposals, see for example, Fowler and Rosenblum, in press) and the arguments for a gestural lexicon should not be construed as supporting this or any other specific proposal. Second, we should note that the use of dynamical equations is not restricted to the description of motor behavior in speech, but has been used to describe the coordination and control of skilled motor actions in general (Cooke, 1980; Kelso, Holt, Rubin, & Kugler, 1981; Kelso & Tuller, 1984a,b; Kugler, Kelso, & Turvey, 1980). Indeed, in its preliminary version, the task dynamic model we are using for speech was exactly the model used for controlling arm movements, with the articulators of the vocal tract simply substituted for those of the arm. Thus, in this respect the model is not consistent with Liberman and Mattingly's (1985) concept of language or speech as a separate module, with principles unrelated to other domains. However, in another respect, the central role of the task in task dynamics captures the same insight as the "domain-specificity" aspect of the modularity hypothesis—the way in which the vocal tract articulators are yoked is crucially affected by the task to be achieved (Abbs, Gracco, & Cole, 1984; Kelso, Tuller, V.-Bateson, & Fowler, 1984).

In this paper, in which we focus on larger language patterns (lexical and phonological structure), we will be demonstrating that many of the patterns observed both in inventories and in alternations can be understood in terms of their articulatory and dynamic structure. The general principles governing motor behavior also interact with other principles, such as articulatory-acoustic relations, in a task-specific way to give rise to a variety of different patterns.

2. LEXICAL ORGANIZATION

In this section, we discuss how gestural structures can fulfill one of the most important functions of lexical representation—distinctiveness. We will examine how contrastive words (or morphemes) differ from one another in terms of their component gestures and their organization. In addition, the same gestural structures have an inherent physical meaning: they directly (without mediation of any "implementation" rules) characterize the articulatory movements of those words.

2.1 Gestures and Distinctiveness

A dynamically defined articulatory gesture characterizes the formation (and release) of a constriction within the vocal tract through the movement of (1) a particular set of articulators, (2) towards a particular constriction location (3) with a specific degree of constriction, and (4) in a characteristic, dynamically-described manner. Gestures are not the movements themselves, but rather abstract characterizations of the movements (see Browman & Goldstein, 1987, and Saltzman et al., 1987, for more detailed descriptions of gestures). Contrastive gestures can be distinguished on the basis of the values of these four attributes.

Because the gestures characterize movements within the vocal tract, they are effectively organized by the anatomy of the vocal tract. For example, at the coarsest level, gestures may differ in the major articulatory subsystems (velic, laryngeal, and oral) employed (attribute (1) above). These choices correspond to contrasts in nasality, voicing, and place. The oral subsystem can be further divided into three distinct articulator sets (or synergies), one for the lips, one for the tongue tip, and one for the tongue body. This (hierarchical) articulatory organization is also incorporated in a number of recent approaches to phonological features (e.g., Clements, 1985; Ladefoged & Halle, 1988; Sagey, 1986). Contrasts in the familiar oral place of articulation involve one of these articulator sets moving to a particular constriction location (attribute (2)).

Gestures may also contrast in the degree of such constriction (attribute (3)), corresponding, for example, to stop-fricative-approximant contrasts. Within our computational model, these three attributes determine the set of "tract variables" (Figure 1). Each oral gesture is represented as a pair of tract variable dynamical equations (one for constriction location, the other for constriction degree), each for the appropriate articulatory synergy. Velic and glottal gestures involve single tract variable equations (Browman & Goldstein, 1987; Saltzman et al., 1987). These dynamical equations include (in addition to specification of target value) dynamic descriptors such as stiffness (related to the rate of movement) and damping. Contrasts in these parameters (attribute (4)) may be relevant, for example, to the distinction between vowels and glides.

Thus, contrasts among gestures involve differences in the attributes (1-4) discussed above. We may ask, however, how the discreteness associated with lexical contrasts emerges from these differences in attributes. First, it should be noted that gestures involving movements of different sets of articulators (attribute 1) differ in an inherently discrete way and are thus automatically candidates for distinctiveness (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The importance of this anatomical source of distinctiveness can be seen in the fact that, for the 317 languages reported in Maddieson (1984), 96.8% have nasals, and about 80% use voicing differences among consonants. Moreover, 98.4% have a contrast among gestures of the lips, tongue tip, and tongue body.

With respect to the other attributes, however, gestures may not differ discretely from an articulatory and/or dynamic perspective (although our knowledge about possible principles governing the control and coordination of speech gestures is, at present, quite limited). In these cases, other forces will tend to create discreteness (and hence the automatic potential for distinctive use). For example, as Stevens (1972) has argued, the potential continuum of constriction degree can be partitioned into relatively discrete regions that produce complete closure or turbulence, based on aerodynamic considerations. The stability of these regions is reflected in the fact that all the languages discussed in Maddieson (1984) have stops; 93.4% have fricatives. For wider constriction degrees, as well as for constriction location (for a given set of articulators), articulatory-acoustic relations (Stevens, 1972) and the predisposition to maximally differentiate the gestures and sounds of a language (e.g., Lindblom, MacNeilage, & Studdert-Kennedy, 1983) will help to localize the distinctive values of these variables in fairly well-defined, stable regions for any one language. These regions may, in fact, differ from language to language (e.g., Disner, 1983; Ladefoged, 1984; Lindblom, 1986), indicating that there is some nondeterminacy in the distribution of these regions. Thus the distinctiveness of individual gestures is a confluence of general articulatory, acoustic, and perceptual pressures towards discreteness, with a certain amount of arbitrariness resolved in language-particular ways.

2.2 Gestural Organization and Distinctiveness

Gestures capture distinctiveness not only individually but also in their organization with respect to each other. Here the fact that gestures are both spatial and temporal becomes crucial. That is, because gestures are characterizations of spatiotemporal articulatory events,. it is possible for them to overlap temporally in various ways. This fact of overlap, combined with the underlying anatomical structure, gives rise to different organizations that can be used contrastively, and also leads to a variety of phonological processes. The distinctive use of overlap will be discussed in this section, its role in phonological processes in section 3.

tract variable		articulators involved
LP	lip protrusion	upper & lower lips, jaw
LA	lip aperture	upper & lower lips, jaw
TTCL	tongue tip constrict location	tongue tip, body, jaw
TTCD	tongue tip constrict degree	tongue tip, body, jaw
TBCL	tongue body constrict location	tongue body, jaw
TBCD	tongue body constrict degree	tongue body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis

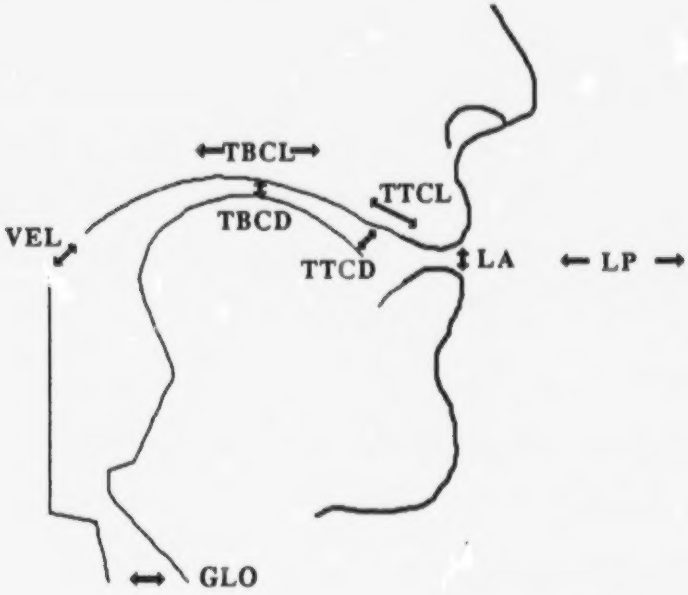


Figure 1. Tract variables and associated articulators used in the computational model of phonology and speech production described in the text.

2.2.1 Gestural Scores

A lexical item typically consists of a characteristic organization of several gestures. Within the computational model being developed, the concept of phase relations (Kelso & Tuller, 1985) is used to coordinate the set of gestures for a given lexical item. This results in a representation that we call a gestural score (Browman et al., 1986). Figure 2a shows a schematic gestural score for the item "spam." The five separate rows can be thought of as "tiers": one each for the velic and glottal subsystems, and three for the oral subsystem, representing the three articulatory synergies. Each box represents a single gesture, its horizontal extent indicating the interval of time during which it is active. (For more details on gestural scores and their use, see Browman & Goldstein, 1987.)

Figure 2 compares two gestural scores, one for "spam" (a) and one for "Sam" (b). There are two additional aspects of distinctiveness that appear when these scores are compared. The only difference between the scores is in the presence of the first bilabial closure gesture in "spam." The first implication of this is that distinctiveness can be conveyed by the presence or absence of gestures (Browman & Goldstein, 1986a; Goldstein & Browman, 1986), which is an inherently discrete property. The second implication is that distinctiveness is a function of an entire constellation of gestures, and that more complex constellations need not correspond to concatenated segment-sized constellations. That is, the difference between "spam" and "Sam" is represented by the difference between two gestures (alveolar fricative and bilabial closure) vs. one gesture (alveolar fricative), in each case co-occurring with a single glottal gesture. This contrasts with a segment based description of the distinction as two segments vs. one segment, where each segment includes both a glottal and an oral constriction specification (see Browman & Goldstein, 1986a, for further discussion). In this sense, the gestural structures are topologically similar to autosegmental structures postulated on the basis of evidence from phonological alternations (e.g., Clements & Keyser, 1983; Hayes, 1986). Note that, in contrast to segment-based theories, which effectively require simultaneous coordination among all the features composing a segment, there are no *a priori* constraints on intergestural organization within the gestural framework (other than anatomical structure). The relative "tightness" of cohesion among particular constellations of gestures is a matter for continuing research.

2.2.2 Gestural Overlap

The gestural scores in Figure 2 show substantial temporal overlap among the various gestures. In this section, we examine how differences in degree of overlap can be used distinctively in the case of two gestures, each with approximately the same extent in time as could occur with some "singleton" consonants, for example syllable-initial [n] (velic and oral gestures) and [t] (glottal and oral gestures). The different possibilities for overlap of such gestures are exemplified in Figure 3, where (as in the gestural scores of Figure 2), the horizontal extent of each rectangle represents the temporal interval during which a particular gesture is active. There is a potential continuum ranging from complete synchrony (displayed in row (a)) through partial overlap (row (b)) to minimal overlap (row (c)). Depending on the particular articulatory subsystems involved (shown in the different columns in the figure), as well as the amount of overlap, these gestural combinations have been categorized (by phonologists and phoneticians) as being very different phenomena, as indicated by the labels for each example in the figure. Various, and in some cases *ad hoc*, phonological features have been employed to capture contrasts in overlap patterns (see Browman & Goldstein, 1986a). We would propose, however, that direct analysis of these organizations in terms of degree of overlap leads to a simpler and more explanatory description of the distribution of these structures in phonological inventories and of their role in phonological processes.

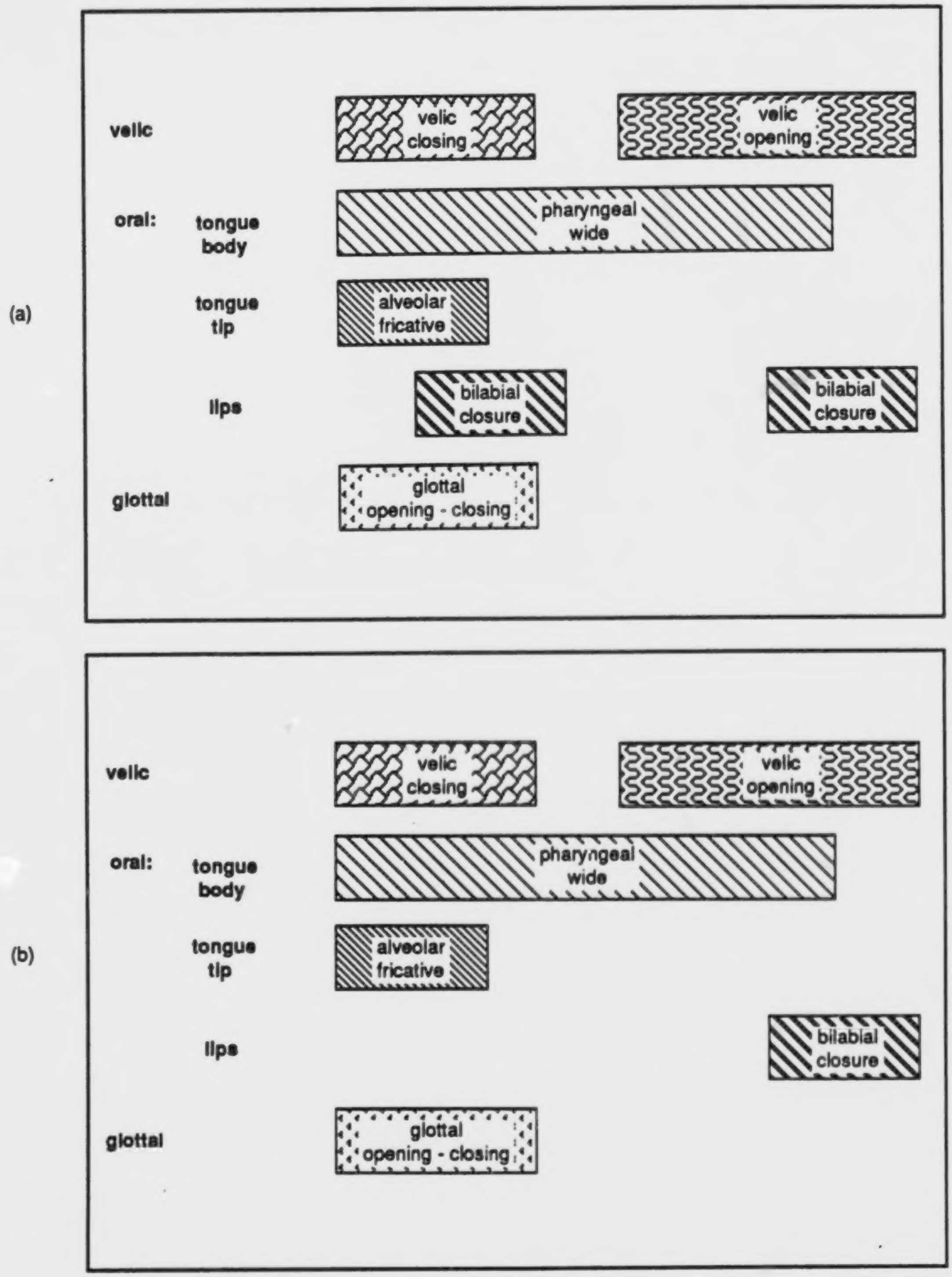

(a)
velic
velic closing
velic opening
oral:
tongue body
pharyngeal wide
tongue tip
alveolar fricative
lips
bilabial closure
bilabial closure
glottal
glottal opening - closing
(b)
velic
velic closing
velic opening
oral:
tongue body
pharyngeal wide
tongue tip
alveolar fricative
lips
bilabial closure
glottal
glottal opening - closing

Figure 2. (a) Gestural score for the word *spam*. (b) Gestural score for the word *Sam*.

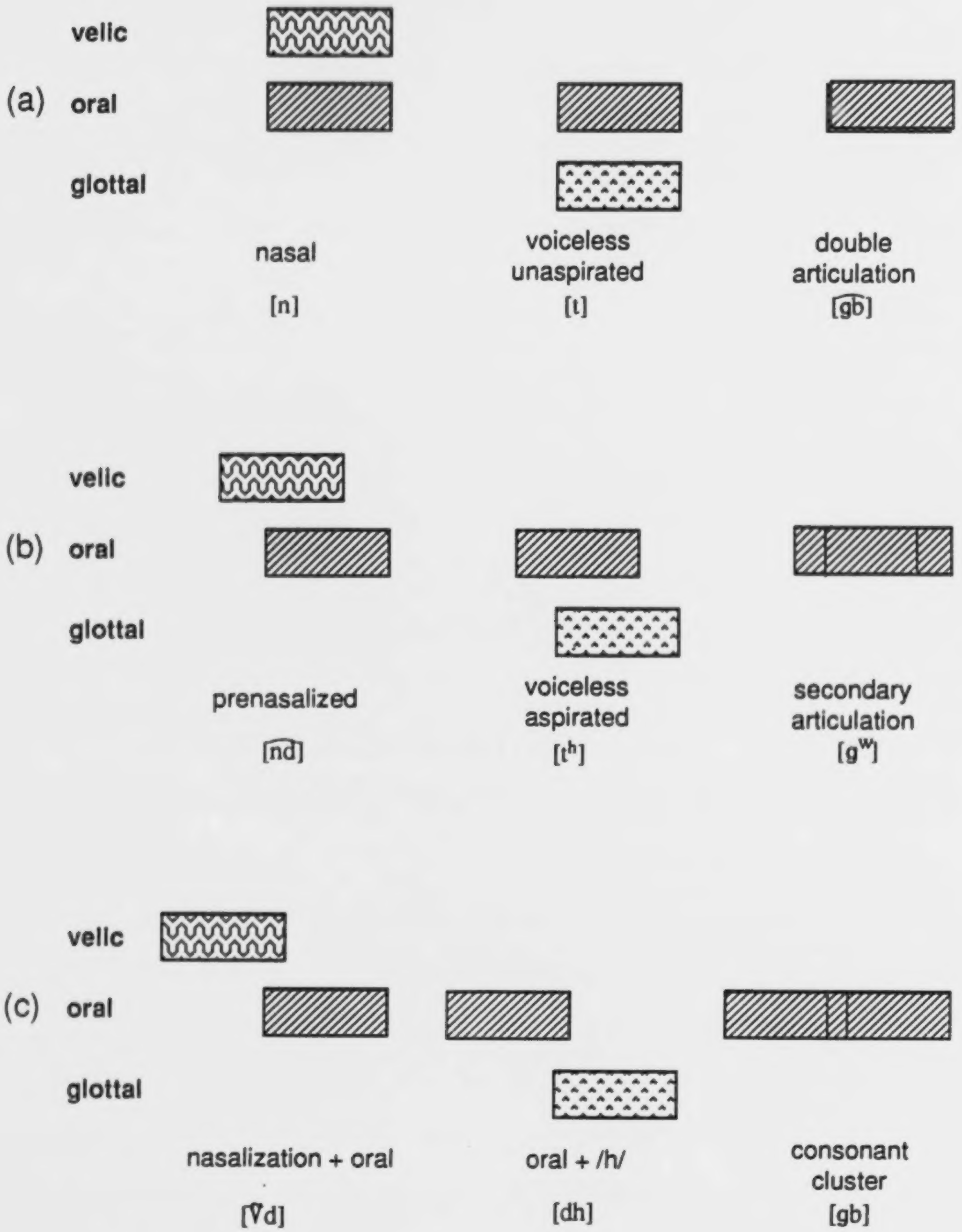

velic
(a)
oral
glottal
nasal
[n]
voiceless
unaspirated
[t]
double
articulation
[g͡b]
velic
(b)
oral
glottal
prenasalized
[n͡d]
voiceless
aspirated
[tʰ]
secondary
articulation
[gʷ]
velic
(c)
oral
glottal
nasalization + oral
[Ṽd]
oral + /h/
[dh]
consonant
cluster
[gb]

Figure 3. Examples of gestural overlap. Each box represents the temporal interval during which a particular gesture is active. (a) Complete synchrony. (b) Partial overlap. (c) Minimal overlap. The leftmost column contrasts degrees of overlap for velic plus oral gestures, the middle column for glottal plus oral gestures, and the rightmost column for two oral gestures.

The simplest case is when two gestures from different articulatory subsystems overlap completely, both having approximately the same extent in time (Figure 3a, left and middle). This case, which (in some sense) maps in a 1-1 fashion onto a segmentation of the acoustic signal, is often analyzed as a segmental unit, for example, (initial) nasal stops or voiceless unaspirated stops (velic plus oral or glottal plus oral, respectively).

What happens to this simple case, however, when one of the gestures is slid slightly so that the overlap is partial rather than complete (Figure 3b, left and middle)? An aspirated, rather than unaspirated stop, for example, maps onto a sequence of two distinct acoustic events—voiceless stop and aspiration—leading to difficulties, in a purely temporal acoustic measure, in assigning the aspiration to the preceding stop or the following vowel. From an articulatory point of view, however, there are still two gestures simply organized slightly differently, such that there is not complete synchrony between the two (see Lisker & Abramson, 1964; Lofqvist, 1980). About 50 languages in Maddieson's (1984) survey contrast these two patterns of organization (voiceless unaspirated vs. voiceless aspirated). The similar overlap differences between the velic and oral gestures are contrastive in 15 of the 18 languages with prenasalized stops, i.e., prenasalized stops contrast with nasals (see also Browman & Goldstein, 1986a; Herbert, 1986).

The status of overlap between two gestures within the same articulatory subsystem (oral-oral: Figure 3, right column) differs in several respects from that of overlap between gestures from different articulatory subsystems (oral-glottal or oral-velic). Overlapping oral (consonant) gestures are relatively rare, and seem to be restricted to two types. First, synchronous double oral articulations (Figure 3a: right) involving stops are possible, although they occur in only 6.3% of the languages surveyed by Maddieson (1984); moreover, they are apparently restricted to labial and velar closure gestures. Relatively synchronous alveolar and velar closure gestures also occur, albeit rarely, but appear to be associated with a different airstream mechanism—velaric suction—that produces clicks (Traill, 1985). (In fact, Marchal [1987] suggests the possibility that the suction associated with clicks could develop automatically from the increased overlap of sequential velar and alveolar closures, as he observes currently in French.) Labial-alveolar double articulations have been reported for such languages as Bura and Margi, but the data of Maddieson (1983) show that the gestures involved display, at most, only partial overlap (but cf. Sagey, 1986, for further discussion of the phonological status of these structures). Second, stops may show at least partial overlap with a gesture of constriction degree wider than stop or fricative (Figure 3b: right). For example, labialization of stops is observed in about 13% of the languages in Maddieson (1984) and roughly the same proportion of languages show palatalization. Other than these cases, the possibilities for overlap of oral closure gestures with other constriction degrees (e.g., fricatives) is controversial (see Sagey, 1986). Thus, while it is clear that languages can have two synchronous oral closure gestures, and can also have two such gestures with minimal overlap (that is, consonant clusters as in Figure 3c: right), it is not clear whether any single language contrasts degrees of overlap for oral stops and fricatives.

Thus, when we examine contrast in terms of gestures and their organization, we find restrictions or "gaps" in phonological inventories. These restrictions involve the distinctive use of overlap among oral gestures (compared to the relatively freer use of overlap contrasts across articulatory subsystems) and can probably be accounted for by a combination of anatomical and acoustic factors. Labials and velars, for example, use different sets of articulators (lips and tongue), so that it is possible to produce them simultaneously (see Halle, 1982); they also have similar, mutually reinforcing effects on the acoustic signal, both having comparatively more energy in the lower rather than upper end of the spectrum (Jakobson, Fant, & Halle, 1969). Ohala and Lorentz (1978) argue that no other combination of constrictions formed with independent articulators shows this type of acoustic compatibility, and that this could account for the predominance of labiovelars among "double" articulations. Constraints on overlapping of gestures with different constriction degrees can also be understood in terms of the aerodynamic or acoustic interaction of the two gestures. It is simply not possible to have the oral tract simultaneously completely closed and also open enough to permit, for example, frication, even if the two constrictions involve different articulator sets and different locations. The acoustic consequences of the oral tract as a whole will be dominated by the narrowest constriction. Note that this points up an inherent difference between the use of two gestures within the same or within different articulatory subsystems. It is perfectly possible, for example, to have the oral tract closed while the nasal tract remains open: the acoustic consequences of the open nasal tract will not be obscured. As Mattingly (1981) points out, this type of conflict may constrain the organization of oral gestures in order to maintain the perceptual recoverability.

More generally, we may ask, as we did in the case of gestural attributes, how the potential continuum of intergestural overlap (particularly when involving different articulatory subsystems) is partitioned into the three discrete contrasting overlap patterns exemplified in Figure 3. While the answer here is quite speculative, again a combination of dynamic articulatory and acoustic factors may be involved. Recent research on bimanual rhythmic movements has demonstrated discrete, stable coordinative modes (in phase vs. out of phase) whose properties can be understood in terms of differential coupling of non-linear oscillators (e.g., Kay, Kelso, Saltzman, & Schöner, 1987; Turvey, Rosenblum, Kugler, & Schmidt, 1986). To the extent that the coordination of different speech gestures can be analyzed in a similar way, it may be possible to discover analogous coupling modes that correspond to contrastive patterns of gestural overlap. In addition, while the overlap between two gestures forms an articulatory continuum, certain qualitative acoustic events may emerge at some critical degree of overlap. For example, as the glottal gesture slides further to the right between rows (a) and (b) of Figure 3 (middle column), at some point perceptible aspiration will be generated. Such emerging acoustic properties could also contribute to the partitioning of the overlap continuum.

In summary, lexical distinctiveness can be represented in gestural structures in three different ways: differences in gestural attributes, presence vs. absence of particular gestures, and differences in gestural overlap. In many cases, the gestural framework provides a natural basis for discrete categorization, particularly when supplemented with additional aerodynamic, acoustic, and perceptual principles. In addition, certain general tendencies found in the phonological inventories of human languages can be rationalized when the elements of these inventories are described as gestural structures.

3. PHONOLOGICAL PROCESSES

The role of gestures and gestural overlap in phonology is not limited to their ability to capture the distinctions among lexical items. Gestural overlap, in particular, can help explain much variability observed during speaking. For example, as argued in the early formulation of the motor theory (Liberman et al., 1967), much coarticulation can be explained in terms of the overlap of vowel and consonant gestures ("parallel transmission"), and this idea has been incorporated into our gesture-based computational model (Browman & Goldstein, 1987; Saltzman et al., 1987). In addition, as will be exemplified below, simple changes in gestural overlap can account for more extreme forms of variation such as apparent segment assimilations, deletions, and insertions (Browman & Goldstein, 1987). The explanatory power of gestural overlap springs from the fact that gestures are abstract characterizations of the actual movements of the articulators, movements that occur in space and over time. When two gestures using the same articulator set co-occur, the movements of the articulators will be affected by both gestures. Even when two co-occurring gestures use different sets of articulators, the nature of their overlap can lead, as we will see, to interesting discontinuous effects on the acoustic signal.

The consequences of overlapping gestures are particularly clear in casual speech (Browman & Goldstein, 1987), where casual speech is defined as that fluent subset of fast speech in which reductions typically occur (Gay, 1981; Lindblom, 1983). Indeed, Browman and Goldstein (1987) have proposed that all variability in casual speech may be due to gestural reduction and/or changes in gestural overlap. The same patterns that occur in casual speech are also observed in "natural" phonological processes (e.g., Sloat, Taylor, & Hoard, 1978) and in many types of historical change (Browman & Goldstein, 1986b; Pagliuca, 1982). In section 3.1, we discuss these common patterns, exemplifying how gestures and gestural overlap can contribute to our understanding of the patterns in each of these areas. In sections 3.2 and 3.3, we discuss synchronic and diachronic patterns (i.e., phonological alternations and historical change) that do not correspond to casual speech processes, showing how in many cases the gestural framework clarifies the nature of the patterns.

3.1 Synchronic and Diachronic Patterns Attributable to Casual Speech Processes

If we assume that a speaker's knowledge of a lexical item includes a specification of its gestures and their organization, it is possible to provide an explanatory account for many (synchronic and diachronic) patterns of phonological and phonetic variation. First, many kinds of variation that have been described by allophonic or low-level phonetic rules can be modeled as the automatic consequence of talking, in which overlapping gestures are produced concurrently. Second, additional kinds of variation can be modeled as consequences of two general principles governing the realization of gestures in casual speech: reduction and increase in overlap. For both of these types of variation, no explicit changes in the talker's representation of the items need to be assumed. In addition, for many kinds of variation that cannot be modeled in this way, we can nevertheless establish a relationship between specific casual speech processes and the development (over historical time) of parallel changes in the talker's representation of the gestural structures themselves (how this might occur is discussed below). Such changes to the gestural structures may either be limited to only some of the environments in which a morpheme occurs, which would be a synchronic alternation that can be described by a gestural phonological rule, or the change may affect an item in all its environments, which would be an example of lexical change in a gestural structure. In this section, then, we will discuss patterns originating in the speech production mechanism. We will look first at reduction of individual gestures, and then at consequences of overlapping gestures, noting different consequences for overlap when gestures employ the same or different sets of articulators. Each type of pattern will be exemplified in casual speech, phonological alternation, and historical (lexical) change.

Brown (1977) discusses a class of weakenings, or *lenitions*, in casual speech in English, where typical examples involve stop consonants weakening to corresponding fricatives: "because" pronounced as [pxəz], "must be" pronounced as [mʌsβɪ]. These changes are *reductions* in the magnitude of individual stop gestures such that there is incomplete closure, and are an instance of the general tendency in some types of fast speech to reduce the movement amplitude (e.g., Lindblom, 1983). Such reductions can occur as regular alternations and therefore be identified as phonological rules in non-casual speech. For example, Spanish voiced stops are pronounced as fricatives when they occur intervocalically: *diccion* ('diction') pronounced with [ð] in *la diccion*; *guerra* ('war') pronounced with [ɣ] in *la guerra* (Sloat et al., 1978). And finally, such reductions may occur not just as alternants but as the sole pronunciation of the word, thereby changing the constriction degree of the gesture in the lexical item. For example, Latin intervocalic /b/ is lenited to a fricative in modern Romance languages, e.g., Latin *habere* ('have'), Italian *avere*, French *avoir* (Lass, 1984).

We may inquire specifically how the articulatory or acoustic output of a casual speech process (in this case reduction) can lead to more permanent or "regular" changes in the gestural structures that underlie the articulatory movement. One possibility is that some speakers become attuned to particular instances of casual speech variation (e.g., reduction in the output magnitude of some gesture), and actually shift (slightly) the value of the constriction degree parameter (CD) for that gesture in that direction. In the stop/fricative cases noted above, reducing the CD of a stop gesture will increase the likelihood that casual speech processes will result in an output that would be categorized as a fricative as opposed to a stop. The greater preponderance of fricative outputs could lead to a further reduction in CD, and, in general, to a systematic drift in this parameter value until a new stable value is reached (the value for a fricative, in this case), and thus, an effective recategorization is achieved. Such stable values would coincide with the discrete parameter ranges discussed in section 2.1. While appeals to drift have been made in other accounts of sound change (e.g., the "allophone" drift of Hockett, 1965), the interaction of this mechanism with the nature of gestural structures increases the range of phenomena to which it may be relevant. In particular, drift may be found not only in the parameters of individual gestures, but also in their overlap, leading to a variety of different articulatory and acoustic consequences. These consequences can be related to phonological processes.

Gestural *overlap* plays a large role in the formation of phonological patterns. (Indeed, it is conceivable that gestural reduction is partly due to the overlap between vowels and consonants, with their opposite requirements for constriction degree.) The consequence of gestural overlap when two gestures involve the same pair of (oral) tract variables (or, as described in Browman & Goldstein, 1987, they occupy the same articulatory tier) follows automatically from the fact that both use the same articulators, leading to blending of the movements of the two. This kind of *assimilation* occurs in the following examples of casual speech (some from British English): *eight things* pronounced as [ɛ⊥ɪ t̪ˈθɪŋz]; *come from* pronounced as [kʌɱfrəm] also; *this year* pronounced as [ˈðɪʃjɪə] (Brown, 1977; Catford, 1977). Such blendings may also occur in the canonical form of words, particularly because the initial consonant and vowel gestures of a syllable may overlap (Browman & Goldstein, 1987; Fowler, 1983; Öhman, 1966). Such canonical blendings are traditionally described phonologically as instances of allophonic variation, e.g., between front and back /k/ in English—*key* vs. *caw* (Ladefoged, 1982)—and can contribute to sound changes: (pre-)Old English *ceap* (presumed to begin with [k]) becoming Modern English *cheap*, but Old English *cuman* (also presumed to begin with [k]) becoming Modern English *come* (Arlotto, 1972).

When overlapping gestures use different tract variables (and therefore different sets of articulators), they do not affect each other's movements but rather both contribute to the overall vocal tract shape, acoustic output, and perceptual response. One consequence of increasing the overlap between oral gestures from different articulatory tiers (i.e., using different tract variables) is (perceptual) assimilation. Various authors (Barry, 1985; Browman & Goldstein, 1987; Kohler, 1976) have presented articulatory evidence that alveolar articulations can occur (possibly reduced in magnitude) in assimilations such as "seven plus" pronounced as [ˈsɛvm̩plʌs] (Browman & Goldstein, 1987). In this example, the following labial gesture overlaps the alveolar gesture to such a degree that the alveolar gesture is effectively "hidden" in the acoustic signal. In addition, the labial gesture also partially overlaps the velic lowering gesture. The result is an apparent assimilation of the place of the nasal to the following consonant (from [n] to [m])—an assimilation that involves no change in the individual gestures, but simply increased overlap among gestures.

Another, more striking, consequence of increasing the overlap between gestures is the percept of *deletion* rather than assimilation. Apparent deletions, which are common in syllable-final position in fluent speech (Brown, 1977), can result from two gestures, on different articulatory tiers, sliding with respect to each other so that one gesture is effectively hidden. Articulatory evidence for sliding has been observed in the utterances "perfect memory" pronounced as [ˈpɚfəkˈmɛməri] and "nabbed most" pronounced as [ˈnaebmows] (Browman & Goldstein, 1987). In these utterances, all the gestures were present: the final alveolar closures in "perfect" and "nabbed" were observed in the movements of the tip of the tongue. However, because of the increased overlap between "perfect" and "memory" (and "nabbed" and "most"), these final alveolar closures were acoustically hidden by the initial labial closures of the following word, resulting in an apparent deletion—acoustically and perceptually, but not articulatorily.

In addition to the casual speech examples above, increased overlap of gestures using different tract variables may also be the source of regular phonological alternations and lexical simplifications. In particular, we hypothesize that an oral gesture may be hidden an increasing proportion of the time through drift in the parameter(s) controlling intergestural phasing. Eventually, a regularly hidden gesture may be deleted from the gestural score, either in particular environments (leading to a synchronic rule), or from the lexical entry entirely. Oral gesture deletion may occur regardless of whether, from a strictly segmental point of view, the perceptual consequences of the increased overlap are partial (i.e., corresponding to casual speech assimilations) or total (i.e., corresponding to apparent deletions) in nature. Examples of both types of gesture deletion can be found, where the likely source of the deletion is, in all cases, hiding due to increased gestural overlap.

An example of oral gesture deletion leading to assimilation involves the assimilation of nasals to the place of the following consonant. Such regular synchronic alternations occur commonly in languages of the world: e.g., Yoruba (an African language) [o ɱ̩ fo] 'he is jumping', [o n̩ lo] 'he is going', [o ŋ̩ ke] 'he is crying' (Bamgbose, 1969, in Sagey, 1986). Nasal place assimilation can also be seen in lexical changes in English, for example in the change from Old French *conforter* to Modern English *comfort* (Arlotto, 1972). Oral gestural deletion of the total type may be seen in phonological rules that delete word-final consonants. For example in Lardil, an Australian language, all non-tongue tip consonants are deleted word-finally: [ŋalu] ('story') vs. [ŋaluk-in] ('story' nonfuture suffix) (Kenstowicz & Kisseberth, 1979). Word-final deletions may also occur in lexical simplifications, for example, Ancient Chinese /fap/ ('law') and /pat/ ('eight') becoming Mandarin /fa/ and /pa/ (Arlotto, 1972).

The hypothesized association between increased overlap and oral gesture deletion can also be seen as underlying the inventory restrictions on the distinctive use of overlapping oral consonantal gestures noted in section 2.2.2. While units consisting of two oral obstruent gestures can arise from overlap, for example in Margi (Hoffman, 1963, in Sagey, 1986), it is clear that there is a strong tendency against such a phenomenon in languages of the world.

Finally, variation in overlap among a whole constellation of gestures can lead to a percept of segment *insertion*. Several authors (e.g., Anderson, 1976; Ohala, 1974) have analyzed the epenthetic stop in nasal-fricative sequences such as *something* [ˊsʌmpθɪŋ] and *Samson* [ˊsaempsən] as arising from variation in the relative timing of the velic closure gesture and the oral closure gesture. In particular, if denasalization precedes the release of the closure gesture, then a short interval of oral closure will be produced. Such variation can lead to historical change, as in the Old English *þymle* becoming Modern English *thimble*. Note that in these cases, no gesture is ever added. Rather, the categorical historical change involves drift in the gestural organization to a different stable pattern of overlap.

3.2 Other Types of Synchronic and Diachronic Patterns

In the previous section, we discussed synchronic and diachronic patterns whose explanation required reference only to the mechanism of speech production, as indicated by their correspondence to casual speech patterns. In this section, we explore some synchronic and diachronic patterns that do not correspond to casual speech variation; that is, they cannot be completely accounted for by the principles of gestural reduction and increase in overlap. A number of these cases have been analyzed as being acoustically or perceptually based. Yet, as we will see, these factors do not appear to operate independently of the articulatory and gestural structures involved, but rather interact with these structures to produce the change (or synchronic alternation). In particular, two additional principles are hypothesized to account for these cases: reassignment of gestural parameters among temporally overlapping gestures (3.2.1) and misparsing of articulatory movements into their underlying discrete gestural regimes (3.2.2).

3.2.1 Reassignment of Gestural Attributes

The ability of listeners to recover the intended gestures from the "single aspect of the acoustic signal" (Liberman et al., 1967, p. 455) that results from gestural overlap is as important in its occasional failure as in its more frequent success. That is, while in the normal course of events, "listeners use coarticulatory information as information for the influencing segment, and ... do not integrate it into their perceptual experience of the segment with which it co-occurs in time" (Fowler, 1984, p. 365), the failure of this ability can lead to changes in the lexical gestural structure (as has been proposed by Ohala, 1981). Such a failure may be exemplified by the change from [x] to [f] in English words such as "cough" and "tough," whose vowels were diphthongs with rounded offglides at the time these changes took place. This sound change has typically been attributed solely to the acoustic similarity of labials and velars (e.g., Jonasson, 1971). However, Pagliuca (1982) shows that most such [x]-[f] changes are not purely acoustic, but rather are conditioned by labial and velar articulations occurring in close proximity prior to the change. That is, in many cases the [x]-[f] change consists of a change in the overlap and constriction degree of the gestures, rather than the insertion of a completely new articulation.

For the English examples like "cough" and "tough," Pagliuca (1982) describes the change as due to "the gradual coarticulation of decaying x with the adjacent rounded diphthong" (p. 171). For words undergoing the [x]-[f] change, the rounded diphthong apparently changed to an unrounded monophthong at the same time that the [x] was changing to [f] (Dobson, 1968, in Pagliuca, 1982). In the gestural framework, all these changes might result simply from an increase in overlap (drifting over time) between the second element of the diphthong ([ʊ]) and the following velar fricative. Figure 4(a) shows the hypothesized gestural score before the change, for the VC portion of the word. Both [o] and [ʊ] describe two gestures: a tongue body gesture (uvular narrow or velar narrow, respectively), and a narrow bilabial gesture for rounding. Increasing overlap, as shown in Figure 4(b), would automatically shorten the preceding vowel from a diphthong to a monophthong. As shown in the figure, it would also mean that the lip-rounding gesture for [u] co-occurred with the frication from the velar gesture. Only one additional step would then be needed: the attribution of the frication (presumably on the part of the listener) to the labial gesture rather than to the tongue body (velar) gesture. This re-assignment of the constriction degree parameter value appropriate for frication would result in a bilabial fricative (and a velar gesture which subsequently is deleted), leaving the structure in 4(c). The monophthongal vowel also lowers at some point to [ɔ], and the bilabial fricative becomes [f], neither of which is shown in the figure.

Ohala (1981) also argues for the importance of (partially) overlapping articulations in historical change, although his emphasis is on the explanatory role of the listener rather than of articulatory overlap, *per se*. For example, Ohala (1981) attributes many examples of assimilation to the (mis)attribution of some acoustic effect to the segment temporally coincident with the effect rather than to segments in the environment: 8th century Tibetan /nub/ 'west' becoming /nu:/ but /lus/ 'body' becoming /ly:/ (Michailovsky, 1975, in Ohala, 1981). In this example, Ohala proposes that, because of coarticulation, /lus/ is auditorily [lys], and therefore the lexical change to a front vowel in /ly:/ results from a re-assignment of the acoustic effects of the overlapping /s/ to the vowel articulation, as the /s/ is deleted. (An analysis along these same lines for similar changes in the history of Lisu is presented in Javkin, 1979). In gestural terms, the situation in Tibetan can be represented as in Figure 5. When the alveolar fricative gesture in (a) is deleted (in (b)), the constriction location of the tongue body gesture is recategorized as palatal, rather than velar, in order to account for effects that were originally due to the overlapping alveolar gesture. As Ohala notes, the acoustic effect on the vowel in [lys] (in (a)) need not result from an articulatory fronting of the tongue body itself—a partially overlapping tongue tip constriction will produce the same auditory effect as a fronted tongue body constriction. This is consistent with the analysis of overlap presented in section 3.1. That is, since the alveolar and vowel gestures are on separate tiers, they interact through the acoustic effect of the gestures occurring simultaneously, rather than through actual blending of constriction targets. Finally, note that the increased length of the vowel in /ly:/ is also automatically accounted for by this analysis. As shown in the figure, the duration of the vowel gestures (tongue body and lips) are assumed to not change (from (a) to (b)), but the part of their duration that is hidden by the overlapping alveolar gesture at the earlier stage is uncovered when the alveolar is deleted. This type of explanation for "compensatory lengthening" phenomena was proposed by Fowler (1983).

Thus, while acoustic and perceptual factors are relevant in accounting for the changes described in this section, the overlapping of two gestures and their interaction are also crucial to the account. Note that these changes do not involve adding articulations that were not there to begin with; rather they involve changes in the parameters of gestures that are already present.

3.2.2 Gestural Misparsing

Other examples of historical change appear to involve introduction of a gesture that was not present at an earlier stage. One such apparent example is the historical introduction of nasalized vowels in words without a nasal consonant in the environment: Sanskrit /sarpa/ ('snake') becoming Hindi /sãp/, Sanskrit /švāsa/ ('breath') becoming Hindi /sãs/ (Ohala & Amador, 1981). As analyzed by Ohala and Amador (1981), such "spontaneous" nasalization is acoustically and perceptually based. The acoustic effects on the vowel of high air flow through the open glottis

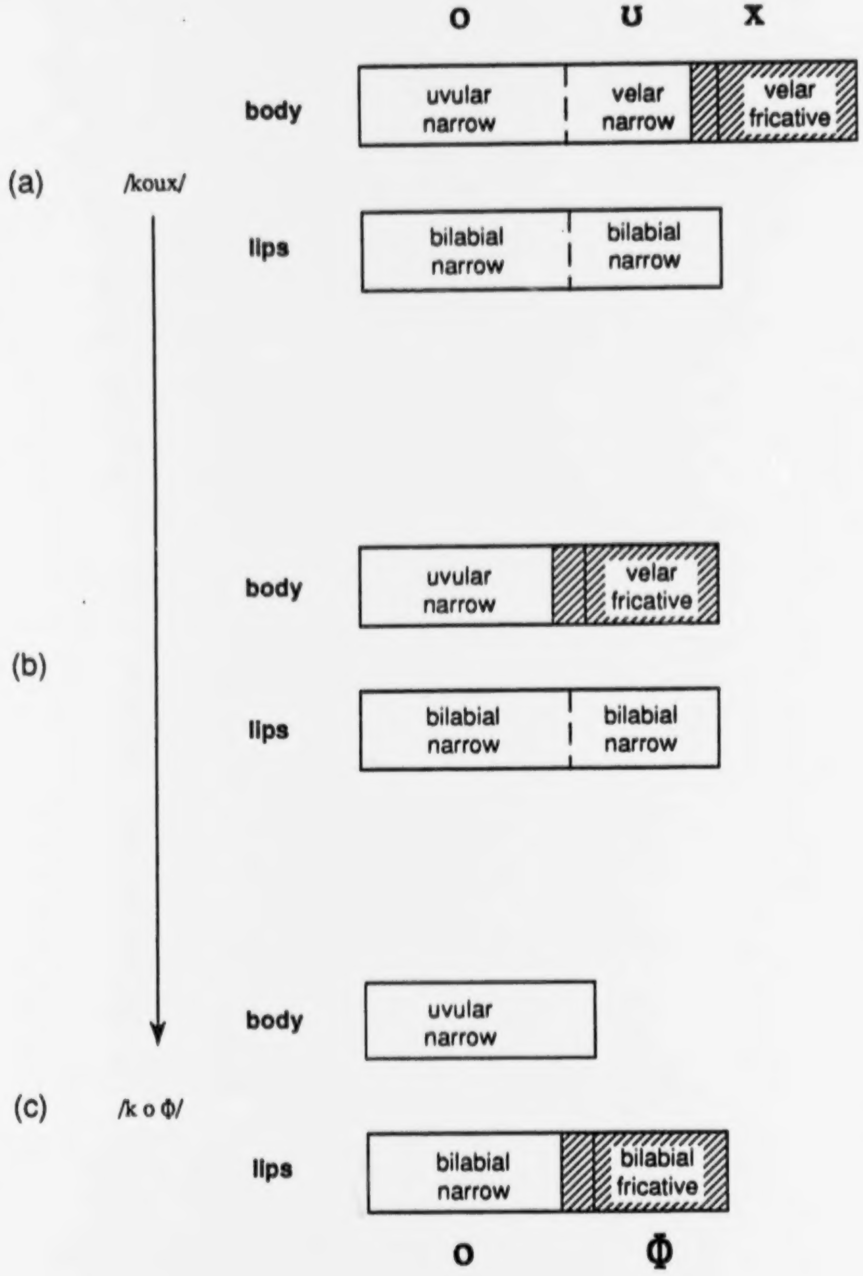

o
u
x
body
uvular narrow
velar narrow
velar fricative
(a)
/koux/
lips
bilabial narrow
bilabial narrow
body
uvular narrow
velar fricative
(b)
lips
bilabial narrow
bilabial narrow
body
uvular narrow
(c)
/k o Φ/
lips
bilabial narrow
bilabial fricative
o
Φ

Figure 4. Hypothesized gestural scores for three stages of sound change for English *cough*. Only the VC portion of the word is shown. (a) is the earliest stage and (c) is the latest.

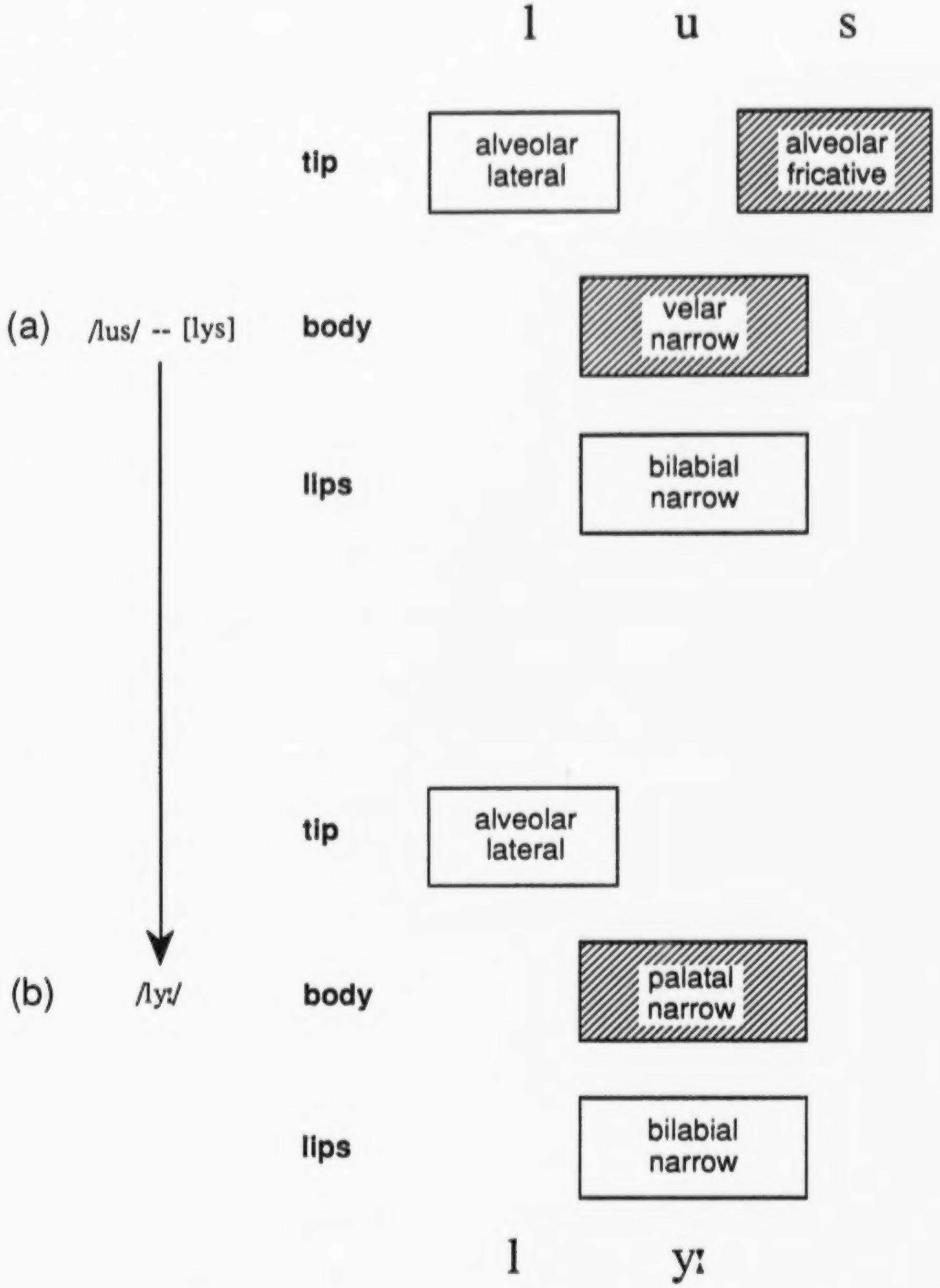

l
u
s
tip
alveolar lateral
alveolar fricative
(a)
/lus/ -- [lys]
body
velar narrow
lips
bilabial narrow
tip
alveolar lateral
(b)
/lyː/
body
palatal narrow
lips
bilabial narrow
l
yː

Figure 5. Hypothesized gestural scores before and after Tibetan sound change: /lus/ — /ly:/.

(especially for fricatives) are re-interpreted by the listener as nasalization, leading to the introduction of a velic opening gesture (in gestural terms). However, there is an alternative (or perhaps complementary) articulatory account suggested by Ignatius Mattingly (personal communication). In general, for oral constriction gestures that are not nasalized, the associated velum height has been found to vary directly with the gesture's constriction degree, decreasing in the series: obstruents—high vowels—low vowels (Bell-Berti, 1980). A relatively low velum position is found during non-nasalized low vowels, even in a language (such as Hindi) that contrasts nasalization on low vowels (Henderson, 1984). Thus, in the normal production of utterances like /sa/, the velum will lower rapidly from the consonant to the vowel of this utterance. This rapid change in velum position may be misinterpreted by a listener (or by a child learning the language) as an explicit velic lowering gesture. This account would be most plausible in the case of low vowel nasalization (as in the cited Hindi examples), and to accept it, we would need to know how often such spontaneous nasalization is associated with low vs. high vowels. To summarize this proposal, if we hypothesize that the continuous articulatory movement (velum lowering) associated with a word is parsed by the language learner into the discrete gestures (and organization) of the gestural score, then this case involves a misparsing: velum lowering is assigned to an explicit gesture. Note however, that in this view, the "new" added gesture is based on an articulation that is already present, rather than being "suddenly" introduced by listeners with no articulatory basis.

This kind of misparsing is very similar to that proposed by Ohala (1981) to account for various kinds of historical dissimilations. An acoustic (or articulatory) effect is attributed to coarticulation with segments in the environment rather than to the segment itself. For example, in the case of pre-Shona /kumwa/ becoming Shona /kumɣa/ (Guthrie 1967-70 and Mkanganwi, 1972, in Ohala, 1981), Ohala proposes that listeners attribute the labiality of the /w/ to the preceding labial consonant, and therefore factor it out, leaving the velar component /ɣ/. Ohala's analysis translated into gestural terms is shown in Figure 6. In (a), the superimposed curve illustrates the waxing and waning of lip constriction (over time) that might be expected when the lips are under the control of the two successive lip gestures shown in the score. In (b) we see that roughly the same lip constriction curve would be expected even if there is only a bilabial closure gesture. As in the nasalization case above, we can view this change in terms of how the observed articulatory movements are parsed into the gestures that could give rise to them. The pattern of lip movement is attributed to a single gesture rather than to a pair of overlapping lip gestures. In this case, the misparsing results in too few gestures (thus deleting one), while in the nasalization case, the misparsing results in a additional gesture.

A related pattern of labial dissimilation occurs synchronically in Cantonese. As described by Yip (1988), labial dissimilation in Cantonese operates on both labial consonants and rounded vowels, that is, on gestures involving the lips. A co-occurrence restriction prevents more than a single gesture using the lips from occurring in the same syllable (except that a back rounded vowel may co-occur with a preceding, but not following, labial consonant). Moreover, this co-occurrence restriction is used productively in a Cantonese "secret language" called La-mi. As discussed in Yip (1988), this secret language uses a form of reduplication in which /C_1V(C_2)/ becomes /lV(C_2) C_1i(C_2)/, e.g., /yat/-→/lat yit/, /kei/-→/lei ki/. However, /sap/ does not become/lap sip/ but rather /lap sit/; similarly, /t'im/ becomes /lim t'in/. That is, the co-occurrence restriction includes the entire two syllables in this secret language, so that a final labial stop is replaced with an alveolar stop in order to preserve the general restriction on labial gestures within a unit. Note that in this secret language, a

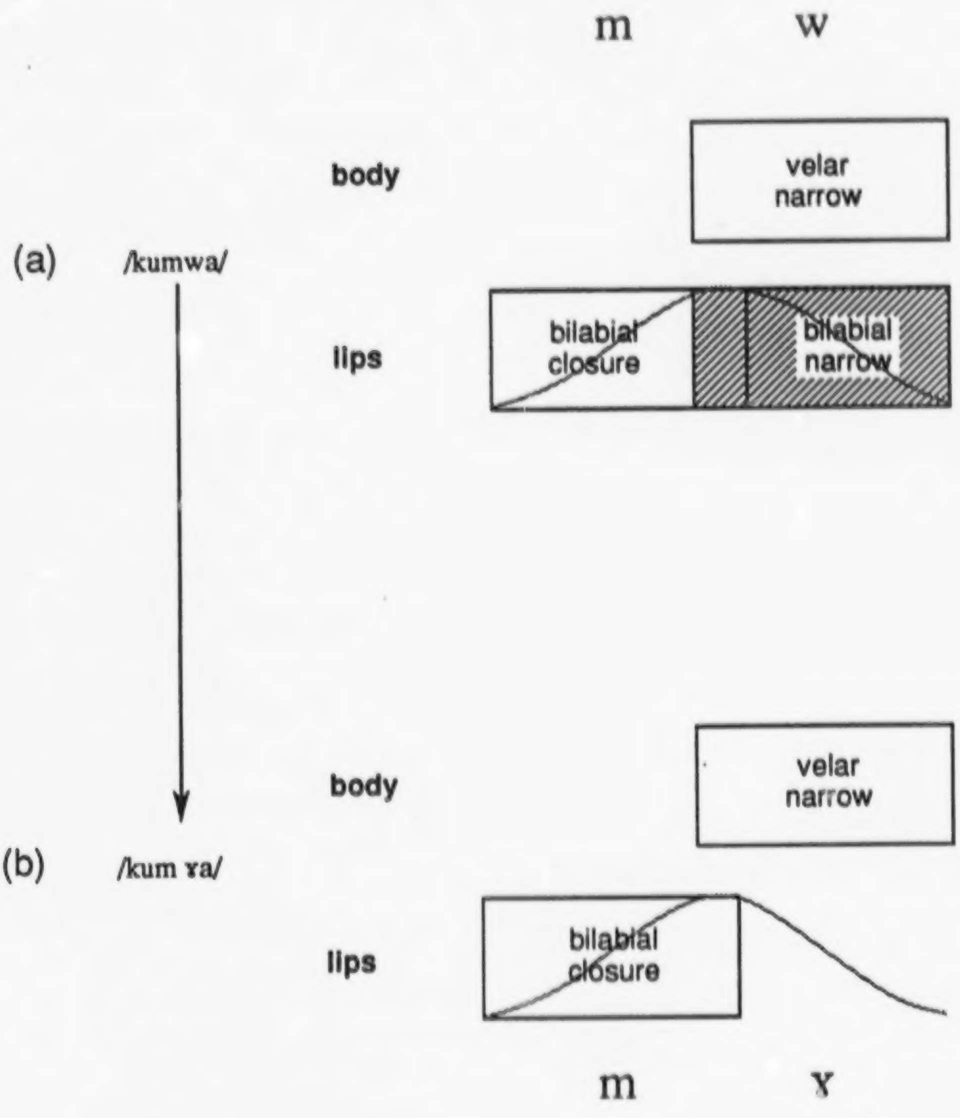

m
w
body
velar
narrow
(a)
/kumwa/
lips
bilabial
closure
bilabial
narrow
body
velar
narrow
(b)
/kum ɣa/
lips
bilabial
closure
m
ɣ

Figure 6. Hypothesized gestural scores before and after Shona sound change: /kumwa/ — /kumɣa/. The curve superimposed on the bilabial gestures shows expected the degree of lip constriction over time resulting from these gestures.

distinctive gesture is neither deleted nor added (as in earlier examples of this section), but rather is replaced by another oral gesture that uses a different set of articulators (i.e., is on a different tier). Moreover, the example shows that a dissimilatory pattern, once set up, can be extended beyond the original misparsed environments that presumably led to the change.

3.3 Patterns Not Attributable to Gestures or Gestural Overlap

Some phonological alternations are so complex as to not permit an adequate description using gestural principles (even with the acoustic and perceptual interactions described in the previous sections). Kenstowicz and Kisseberth (1979) describe a morphophonemic process in Chimwi:ni (a Bantu language spoken in Somalia) that exemplifies the degree of complexity that languages can attain. The process involves an interaction between the perfective suffix for verbs and the final consonant of the stem. The form of the perfective suffix for verbs either contains the voiceless lateral alveolar fricative [ɬ], or, if the stem-final consonant is [s,z,ʃ,ñ], it contains a [z]. This appears to be a kind of assimilatory process. But the class of [s,z,ʃ,ñ] is not a natural class articulatorily: while they are all central, and all use the front part of the tongue, other central tongue tip consonants ([d], [n]) occur with [ɬ] rather than [z].

The behavior of the stem-final consonant adds to the eye-glazing complexity of the perfective form. Stem-final consonants that are stops in the infinitive form generally correspond to the central alveolar fricatives [s] and [z] in the perfective form. That is, the oral gesture can use different articulators in the two forms (for example, [p]/[s]), as well as having a different constriction degree (voicing is unchanged). However, [k] corresponds to [ʃ] not [s], thereby showing some slight hint of articulatory conditioning, either of the specific articulator or front/back constriction location.

While the Chimwi:ni example displays a certain amount of articulatory patterning, it is not possible to provide a general statement of the patterns contributing to the alternations, even using highly abstract projections of gestures (or their acoustic consequences). At some point in every language, the patterns begin to take on a life of their own, loosened from their articulatory and/or acoustic underpinnings, and perhaps respecting other sets of principles (cf. Anderson, 1981). Phonologists have attempted to describe these patterns (e.g, "crazy rules," Bach & Harms, 1972) as emerging from the interaction of a number of independent rules, each of which, by itself, can be simply understood. (Rule telescopy [Wang, 1969] and rule inverson [Vennemann, 1972] are examples of such interaction).

In summary, we have attempted to show how lexical representations can be viewed as gestural structures, and to show that to view them thus contributes to an understanding of phonological inventories and processes. While other principles and sources of constraint are no doubt required to completely explicate patterns in phonology, we have been surprised by the range of phenomena that can be handled with the relatively simple assumptions that we are making: that phonological structures consist of temporally overlapping, dynamically defined gestural units, that the output of these structures may be systematically altered (in highly constrained ways) in casual speech, and that variations in output of these structures can lead (historically) to changes in the values of gestural parameters—both through drift and through mechanisms such as re-assignment and misparsing. Indeed, several of the cases that we present in this paper were initially chosen by us to illustrate where the gestural approach would fail. Once the gestural analyses were made explicit, however, they were more insightful than even we expected. We suggest that it is interesting and important to see just how much such simple structures and principles can contribute to understanding phonological patterns.

ACKNOWLEDGMENT

This work was supported by NSF grant BNS 8520709 and NIH grants HD-01994 and NS-13617 to Haskins Laboratories. Our thanks to Margaret Dunn, Alice Faber, Ignatius Mattingly, Elliot Saltzman, and Douglas Whalen for helpful comments.

REFERENCES

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multimovement coordination: Sensorimotor mechanisms in speech motor programming. *Journal of Motor Behavior, 16,* 195-231.

Anderson, S. R. (1976). Nasal consonants and the internal structure of segments. *Language, 52,* 326-344.

Anderson, S. R. (1981). Why phonology isn't "Natural." *Linguistic Inquiry, 12,* 493-540.

Arlotto, A. (1972). *Introduction to historical linguistics.* Boston: Houghton Mifflin Co.

Bach, E., & Harms, R. T. (1972). How do languages get crazy rules? In R. P. Stockwell & R. K. S. Macaulay (Eds.), *Linguistic change and generative theory* (pp. 1-21). Bloomington, IN: Indiana University Press.

Bamgbose, A. (1969). Yoruba. In E. Dunstan (Ed.), *Twelve Nigerian languages.* New York: African Publishing Corp.

Barry, M. C. (1985). A palatographic study of connected speech processes. *Cambridge Papers in Phonetics and Experimental Linguistics, 4,* 1-16.

Bell-Berti, F. (1980). Velopharyngeal function: a spatio-temporal model. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice, Vol. IV,* (pp. 291-316). New York: Academic Press.

Browman, C. P., & Goldstein, L. (1986a). Towards an articulatory phonology. *Phonology Yearbook, 3,* 219-252.

Browman, C. P., & Goldstein, L. (1986b). Dynamic processes in linguistics: Casual speech and historical change. *PAW Review, 1,* 17-18.

Browman, C. P., & Goldstein, L. (1987). Tiers in Articulatory Phonology, with some implications for casual speech. *Haskins Laboratories Status Report on Speech Research, SR-92,* 1-30. To appear in J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech.* Cambridge: Cambridge University Press.

Browman, C. P., Goldstein, L., Kelso, J. A. S., Rubin, P., & Saltzman, E. (1984). Articulatory synthesis from underlying dynamics. *Journal of the Acoustical Society of America, 75,* S22-S23. (Abstract)

Browman, C. P., Goldstein, L., Saltzman, E., & Smith, C. (1986). GEST: A computational model for speech production using dynamically defined articulatory gestures. *Journal of the Acoustical Society of America, 80,* S97. (Abstract)

Brown, G. (1977). *Listening to spoken English.* London: Longman.

Catford, J. C. (1977). *Fundamental problems in phonetics.* Bloomington, IN: Indiana University Press.

Clements, G. N. (1985) The geometry of phonological features. *Phonology Yearbook, 2,* 225-252.

Clements, G. N., & Keyser, S. J. (1983). *CV phonology: A generative theory of the syllable.* Cambridge, MA: MIT Press.

Cooke, J. D. (1980). The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 199-212). Amsterdam: North-Holland.

Disner, S. F. (1983). Vowel quality: The relation between universal and language specific factors. *UCLA Working Papers in Phonetics, 58,* Los Angeles, CA.

Dobson, E. J. (1968). *English pronunciation 1500-1700* (2 vols., 2nd. ed.). Oxford: Clarendon Press.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General, 112,* 386-412.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics, 36,* 359-368.

Fowler, C. A., Rubin, P., Remez, R.E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production* (pp. 373-420). New York: Academic Press.

Fowler, C. A., & Rosenblum, L. (in press). The perception of phonetic gestures. In I. G. Mattingly and M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Erlbaum Associates.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica, 38*, 148-158.

Goldstein, L., & Browman, C. P. (1986). Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics, 14*, 339-342.

Guthrie, M. (1967-70). *Comparative Bantu* (4 vols). Farmborough: Gregg.

Halle, M. (1982). On distinctive features and their articulatory implementation. *Natural language and linguistic theory, 1*, 91-105.

Hayes, B. (1986). Assimilation as spreading in Toba Batak. *Linguistic Inquiry, 17*, 467-500.

Henderson, J. (1984). *Velopharyngeal function in oral and nasal vowels: A cross-language study*. Doctoral dissertation, University of Connecticut.

Herbert, R. K. (1986). *Language universals, markedness theory, and natural phonetic processes*. Berlin: Mouton de Gruyter.

Hockett, C. F. (1965). Sound change. *Language, 41*, 185-204.

Hoffman, C. (1963). *A grammar of the Margi language*. London: Oxford University Press.

Jakobson, R., Fant, C. G. M., & Halle, M. (1969). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge: The MIT Press.

Javkin, H. (1979). Phonetic universals and phonological change. *Report of the Phonology Laboratory, 4*, University of California, Berkeley.

Jonasson, J. (1971). Perceptual similarity and articulatory reinterpretation as a source of phonological innovation. *STL-QPSR, 1*, 30-42.

Kay, B. A, Kelso, J. A. S., Saltzman, E. L., & Schöner, G. (1987). Space-time behavior of single and bimanual rhythmical movements: Data and limit cycle model. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 178-192.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data. *Journal of Motor Behavior, 13*, 226-221.

Kelso, J. A. S., & Tuller, B. (1984a). A dynamical basis for action systems. In M. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum. 321-356.

Kelso, J. A. S., & Tuller, B. (1984b). Converging evidence in support of common dynamical principles for speech and movement coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 246*, R928-R935.

Kelso, J. A. S., & Tuller, B. (1985). Intrinsic time in speech production: Theory, methodology, and preliminary observations. *Haskins Laboratories Status Report on Speech Research, SR-81*, 23-39.

Kelso, J.A.S., B. Tuller, Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 812-832.

Kenstowicz, M., & Kisseberth, C. (1979). *Generative phonology*. Orlando, Fla: Academic Press.

Kohler, K. (1976). Die Instabilitat wortfinaler Alveolarplosive im Deutschen: eine elektropalatographische Untersuchung [The instability of word-final alveolar plosives in German: An electropalatographic investigation.] *Phonetica, 33*, 1-30.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 3-47). New York: North-Holland.

Ladefoged, P. (1982). *A Course in Phonetics* (2nd ed.). New York: Harcourt Brace Jovanovich.

Ladefoged, P. (1984). 'Out of chaos comes order': Physical, biological, and structural patterns in phonetics. In M. P. R. van den Broecke & A. Cohen (Eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences* (pp. 83-95). Dordrecht, Holland: Foris Publications.

Ladefoged, P., & Halle, M. (1988). Some major features of the International Phonetic Alphabet. *Language, 64*, 577-582.

Lass, R. (1984). *Phonology*. Cambridge: Cambridge University Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-436.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech*. New York: Springer-Verlag.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology*. Orlando, FL: Academic Press.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations of Linguistic Universals*. The Hague: Mouton.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384-422.

Löfqvist, A. (1980). Interarticulator programming in stop production. *Journal of Phonetics, 8*, 475-490.

Maddieson, I. (1983). The analysis of complex phonetic elements in Bura and the syllable. *Studies in African Linguistics, 14*, 285-310.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Marchal, A. (1987). Des clics en français. *Phonetica, 44*, 30-37.

Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 415-420). Amsterdam: North-Holland.

Michailovsky, B. (1975). On some Tibeto-Burman sound changes. *Proceedings of the Annual Meeting, Berkeley Linguistics Society, 1*, 322-332.

Mkanganwi, K. G. (1972). The relationships of coastal Ndau to the Shona dialects of the interior. *African Studies, 31*, 111-137.

Ohala, J. J. (1974). Experimental historical phonology. In J. M. Anderson & C. Jones (Eds.), *Historical linguistics II: Theory and description in phonology* (pp. 353-389). Amsterdam: North-Holland.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178-203). Chicago: Chicago Linguistic Society.

Ohala, J. J., & Amador, M. (1981). Spontaneous nasalization. *Journal of the Acoustical Society of America, 69*, S54-S55. (Abstract)

Ohala, J. J., & Lorentz, J. (1978). The story of [w]: An exercise in the phonetic explanation for sound patterns. *Report of the Phonology Laboratory, 2*, 133-155.

Ohman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39*, 151-168.

Pagliuca, W. (1982). *Prolegomena to a theory of articulatory evolution*. Doctoral dissertation, SUNY at Buffalo. Ann Arbor: University Microfilms International.

Sagey, E. C. (1986). *The representation of features and relations in non-linear phonology*. Unpublished doctoral dissertation, MIT.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns (Experimental Brain Research Series 15)*. New York: Springer-Verlag.

Saltzman, E., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. *Psychological Review, 94*, 84-106.

Saltzman, E., Rubin, P. E., Goldstein, L. & Browman, C. P. (1987). Task-dynamic modeling of interarticulator coordination. *Journal of the Acoustical Society of America, 82*, S15. (Abstract)

Sloat, C., Taylor, S. H., & Hoard, J. E. (1978). *Introduction to phonology*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51-66). New York: McGraw Hill.

Traill, A. (1985). *Phonetic and phonological studies of !XOO Bushman*. Hamburg: Helmut Buske Verlag.

Turvey, M. T., Rosenblum, L. D., Kugler, P. N., & Schmidt, R. C. (1986). Fluctuations and phase symmetry in coordinated rhythmic movements. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 564-583.

Vennemann, T. (1972). Rule inversion. *Lingua*, 209-242.

Wang, W. (1969). Competing changes as a cause of residue. *Language, 45*, 9-25.

Yip, M. (1988). The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry, 19:1*, 65-100.

FOOTNOTES

[*] To be published in I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Erlbaum Associates.

[†] Also Yale University.

The Cricothyroid Muscle in Voicing Control*

Anders Löfqvist,[†] Thomas Baer, Nancy S. McGarr,[††] and Robin Seider Story[†††]

Initiation and maintenance of vibrations of the vocal folds require suitable conditions of adduction, longitudinal tension, and transglottal airflow. Thus, manipulation of adduction/abduction, stiffening/slackening, or degree of transglottal flow may, in principle, be used to determine the voicing status of a speech segment. This study explores the control of voicing and voicelessness in speech with particular reference to the role of changes in the longitudinal tension of the vocal folds, as indicated by cricothyroid (CT) muscle activity. Electromyographic recordings were made from the CT muscle in two speakers of American English and one speaker of Dutch. The linguistic material consisted of reiterant speech made up of CV syllables where the consonants were voiced and voiceless stops, fricatives, and affricates. Comparison of CT activity associated with the voiced and voiceless consonants indicated a higher level for the voiceless consonants than for their voiced cognates. Measurements of the fundamental frequency (F0) at the beginning of a vowel following the consonant show the common pattern of higher F0 after voiceless consonants. For one subject, there was no difference in cricothyroid activity for voiced and voiceless affricates; in this case, the consonant-induced variations in the F0 of the following vowel were also less robust. Consideration of timing relationships between the EMG curves for voiced and voiceless consonants suggests that the differences most likely reflect control of vocal-fold tension for maintenance or suppression of phonatory vibrations. The same mechanism also seems to contribute to the well-known difference in F0 at the beginning of vowels following voiced and voiceless consonants.

INTRODUCTION

For sustained vibrations of the vocal folds to occur, some physiological and aerodynamic conditions must be met (cf. Stevens, 1977; Titze, 1980; van den Berg, 1958). In the larynx, the vocal folds must be adducted to a suitable degree and the longitudinal tension of the folds must be adjusted within an appropriate range. In addition, a transglottal air flow is required. This flow is induced by a transglottal pressure drop and it in turn induces variations in pressure within the glottis, depending on the shape and configuration of the vocal folds. When conditions for phonation are met, these variations in pressure and the movements of the vocal folds that they cause interact with each other to produce sustained vibrations (cf. Titze, 1986).

Manipulation of the conditions that are necessary for vibrations to occur can, in principle, also be used to control the cessation of voicing during periods of voicelessness in speech. Thus adduction/abduction and stiffening/slackening of the vocal folds and facilitation/inhibition of transglottal flow are all potential mechanisms for determining the voicing status of a speech segment. The roles of adduction/abduction and facilitation/inhibition of air-flow are relatively well understood. However, there is no general agreement whether control of longitudinal tension of the folds is actually used to control voicing status. The present work is aimed at clarifying the role of this potential mechanism.

Vocal-fold abduction is the mechanism most often associated with voicelessness in speech. The glottal abduction gesture is phased with respect to oral articulations to produce contrasts of aspiration, and its amplitude and duration also vary with consonant type (e.g., Löfqvist, 1980; Löfqvist & McGarr, 1987: Löfqvist & Yoshioka, 1980, 1984). However, glottal abduction in itself may not always be sufficient for preventing laryngeal vibrations.

Glottal abduction during speech production is usually combined with a decrease of the air flow through the glottis due to a supraglottal constriction. For voiceless stop consonants, there is a build-up of oral air pressure behind the supraglottal occlusion and the pressure drop across the glottis is thus decreased. The glottal vibrations cease as the transglottal flow decreases and the glottal opening increases (cf. Hirose & Niimi, 1987; Yoshioka, 1984). Venting the vocal tract during the period of stop closure, thus restoring the transglottal flow, may sometimes result in glottal vibrations (Perkell, 1976; Vencov, 1968). For voiceless fricatives, where the supraglottal constriction is incomplete, there is greater airflow than for stops, but the decrease in transglottal pressure drop in combination with the glottal abduction gesture may cause cessation of glottal vibrations. On the other hand, for the laryngeal fricative /h/, an abduction gesture is often found but the vibrations may still continue; in this case, no supraglottal constriction is made and the transglottal flow continues uninterruptedly. Thus, the laryngeal abduction gesture may not in itself arrest glottal vibrations but may do so only in combination with aerodynamic factors (cf. Yoshioka, 1981). The interaction between glottal abduction and aerodynamic factors results in a subtle difference between the voicing offsets of voiceless stops and fricatives. Vibrations generally end at a larger glottal opening for fricatives than for stops (Hirose & Niimi, 1987). This fact could be related to the different aerodynamic conditions prevailing at the onset of stop closure and fricative constriction, respectively. In the former case, the vocal tract is completely sealed off, and the air flow through the glottis decreases abruptly. For fricatives, on the other hand, transglottal flow continues throughout the period of oral constriction even though the oral pressure is increased behind the constriction. Even during stop occlusion, aerodynamic factors may be used to facilitate or inhibit glottal airflow, thus contributing to the voicing distinction. This is accomplished through control of supraglottal volume. Active or passive expansion of the vocal tract facilitates voicing while tensing it or actively constricting it inhibits voicing (cf. Bell-Berti, 1975; Minifie, Abbs, Tarlow, & Kwaterski, 1974; Westbury, 1983).

Muscular control of the abduction gesture for devoicing involves suppression of activity in the interarytenoid, lateral cricoarytenoid, and thyroarytenoid muscles, and activation of the posterior cricoarytenoid. In comparison, for (unaspirated) voiced consonants, there may be some suppression of the adductors, but the posterior cricoarytenoid is not active.

While glottal abduction is the laryngeal gesture commonly associated with an arrest of glottal vibrations, its opposite, increased glottal adduction, can also be used for the same purpose. In this case, a tight closure is produced, preventing transglottal flow. Such an adjustment occurs for glottal stops and also for voiceless ejective and ingressive consonants (cf. Fre Woldu, 1985; Hirose & Gay, 1973). Electromyographic recordings of intrinsic laryngeal muscles suggest that the thyroarytenoid muscle is active in this case, adducting the folds and increasing the tension of their bodies.

While changes in glottal abduction/adduction and transglottal flow are certainly used for arresting glottal vibrations, it is more uncertain if the remaining possible mechanism—increased longitudinal tension—is used in speech for the same purpose. This mechanism was suggested by Halle and Stevens (1971) on theoretical grounds, but the evidence supporting it has been equivocal.

Physiological studies have shown that the cricothyroid muscle is most consistently associated with changes in the longitudinal tension of the vocal folds. This muscle controls the length and tension of the vocal folds by rotating the cricoid and thyroid cartilages relative to each other (e.g., Sonesson, 1982). Changes in vocal-fold tension result in changes of fundamental frequency during phonation, and thus cricothyroid muscle activity is highly correlated with fundamental frequency.

Another muscle that could also be involved in the control of the tension of the vocal folds is the vocalis muscle. EMG recordings from the vocalis indicate that it is active for both adduction of the vocal folds and changes in tension (cf. Löfqvist, McGarr, & Honda, 1984); in particular, it adjusts the relative tensions of the vocal fold body and cover. Although the experimental evidence is unclear, the vocalis most often shows reduced activity for voiceless consonants.

While the experimental results are conflicting, some EMG studies have suggested that the activity of the cricothyroid muscle is higher for voiceless than for voiced consonants. Dixit and MacNeilage (1980) reported that the cricothyroid was more active during voiceless than during voiced stops, fricatives, and affricates in Hindi. Löfqvist, et al. (1984) showed that the cricothyroid was more active just before the beginning of the oral closure/constriction for voiceless as opposed to voiced consonants in Swedish. Other studies reviewed by Sawashima and Hirose (1983) also suggest a similar difference. However, Hirose and Ushijima (1978) suggested that these observed differences in CT level could have been associated with differences in the intonation patterns for the utterances involved. Several other investigators report no difference in the CT activity level associated with voiced-unvoiced pairs (e.g., Collier, Lisker, Hirose, & Ushijima, 1979; Kagaya & Hirose, 1975). If such a relative increase in cricothyroid activity for voiceless consonants occurs, it could be related to the well-known higher fundamental frequency of a vowel following a voiceless consonant as compared to a vowel following a voiced consonant (e.g., Hombert, Ohala, & Ewan, 1979; House & Fairbanks, 1953; Lehiste & Peterson, 1961; Ohde, 1984). This phenomenon served as the impetus for hypothesizing a role for laryngeal tension in the voicing distinction. However, the lack of experimental evidence showing a systematic difference in CT activity has led investigators to seek alternate explanations for this phenomenon, such as articulatory and aerodynamic explanations.

There is thus conflicting evidence concerning the role of variations in the longitudinal tension of the vocal folds in the control of voicelessness. The present study was therefore designed to investigate further the activity of the cricothyroid muscle in voicing control using electromyographic techniques.

I. METHODS

Electromyographic recordings were obtained from the cricothyroid muscle in three subjects. Two of the subjects, one male (TB) and one female (NSM), were native speakers of American English; the third subject (LB), was a male native speaker of Dutch.

The linguistic material consisted of reiterant speech modeled after the sentence "The man went to market." (The underline indicates main lexical stress). Different CV syllables, drawn from 16 consonants and 3 vowels, were reiterated to form 48 different utterance types. The vowels used for the reiterant syllables were /i ɑ u/ and the consonants were voiced and voiceless stops (/p b t d k g/), fricatives (/f v θ ð s z ʃ ʒ/), and affricates (/tʃ dʒ/). Only stops and fricatives were used for the Dutch subject. For the analysis of voicing contrasts, we used the consonant occurring in the syllable carrying the sentence stress in the utterance, that is, the second syllable in the sentence. Five to ten repetitions of each utterance type were obtained.

Hooked-wire electrodes were inserted percutaneously under topical anesthesia of the skin. Verification of electrode position was made by having the subject perform selected speech and nonspeech gestures, such as pitch changes, vocal attacks, jaw movements, and swallowing. Placement into the CT was verified by showing activity correlated with F0 but not with adduction, abduction, or jaw opening, and a characteristic pattern of activation and suppression during swallow. After amplification and high-pass filtering at 80 Hz, the signals were recorded on an FM tape recorder together with the speech signal. For averaging, the signals were full-wave rectified, integrated over a 5-ms window, then sampled and digitized, resulting in a computer sampling rate of 200 Hz. In the averaging process, the signals were aligned with reference to a predetermined, acoustically defined line-up point, and also further smoothed using a 35-ms triangular window. The line-up point for fricatives and affricates was the apparent end of frication. The line-up point for stops was the release burst. This point was chosen because it clearly marks the end of the period in which differing activity would be expected, if it were associated with the voicing distinction. Alternatively,. we could have chosen the point of stop closure or onset of frication. However, in the present study, we have chosen to concentrate more on the CV rather than the VC junction.

In order to obtain a single numerical value for the level of cricothyroid activity, the area under the EMG curve for each token was calculated during a short time window. This window was chosen so as to cover the period of consonant occlusion and part of the preceding vowel; its end always coincided with the line-up point. The duration of the window was chosen individually for each subject and consonant type based on acoustic measurements. Typical values were 125 ms for stops, 175 ms for fricatives, and 150 ms for affricates. The position of the window was based on theoretical and empirical considerations. If an increase in the tension of the vocal folds is used to suppress glottal vibrations, the neuromuscular events ought to occur within this window. They would be expected to begin near the transition from the voiced vowel to the voiceless consonant. This is also the point at which Löfqvist et al. (1984) found an increase in cricothyroid activity for voiceless consonants.

Measurements were also made of the fundamental frequency at the beginning of the vowel following the consonant. For this, the duration of the first 10 pitch periods of each token were measured interactively on a computer. To keep the measurement task manageable, these measurements were only made for utterances containing the vowel /ɑ/.

II. RESULTS

A. Electromyography

The averaged electromyographic curves for each of the three subjects are shown in Figures 1-3; in these averages, the utterances containing different vowels have been pooled, since there were no systematic differences across vowels. For the present analysis and display purposes, data have also been pooled across place of articulation for the stops and fricatives. Average audio amplitude envelopes are also shown in Figures 1-3. The audio envelope curves provide a rough indication of consonant/vowel boundaries, though it must be remembered that the tokens contributing to these averages displayed a range of temporal characteristics so that alignment of segment boundaries is imprecise except near the line-up point.

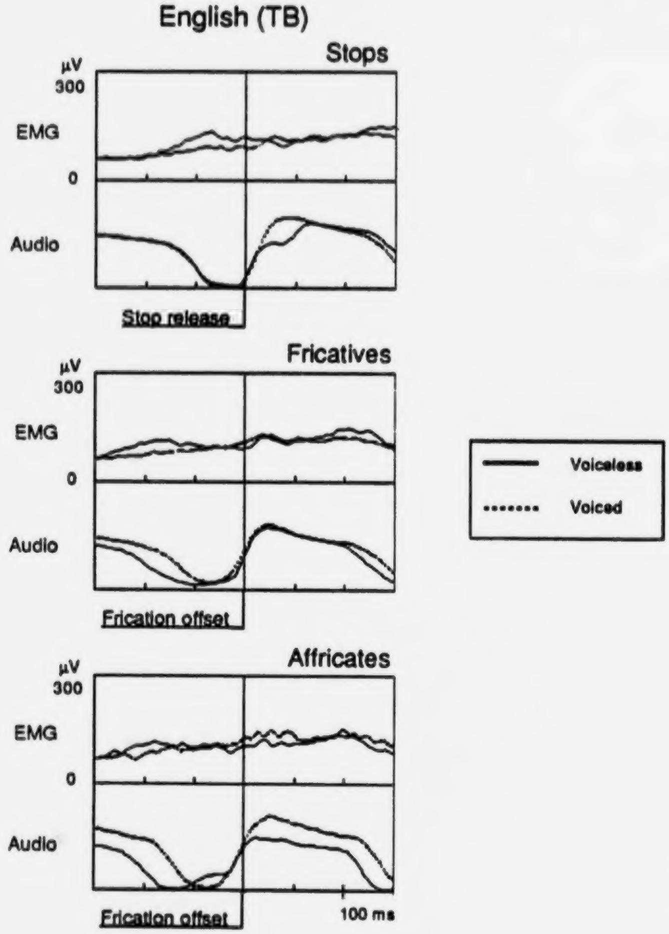


Figure 1. Average EMG and audio signals for subject TB. (n = 47 for the stops and fricatives, and 15 for the affricates.)

The plots of CT activity shown in Figures 1-3 all share a common feature. In the left half of each figure, covering a period associated with the preceding vowel, the transition, and the consonant occlusion, the curves for voiced and voiceless consonants begin at the same level, diverge from each other with the voiceless one showing higher activity, and then reconverge in the vicinity of the line-up point. Qualitatively, where the curves have separated, they are clearly at visually distinct levels in all cases, except, perhaps, for the affricates of subject TB, shown in Figure 1. The divergence of the two curves is due mostly to an increase of activity for the voiceless consonants; activity for the voiced consonants remains roughly constant, increasing later, so that when the curves reconverge at the consonant release, they are at a higher level than for the preceding vowel, in accordance with the stress pattern of the utterance. The curves also appear to diverge in the right halves of each figure, for the following consonant, although the timing is more variable.

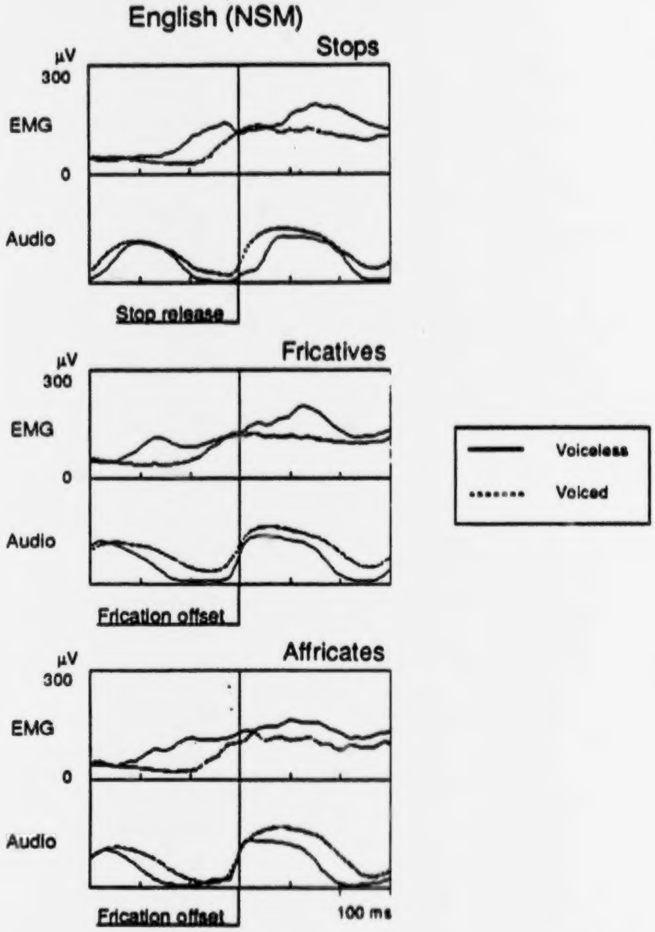


Figure 2. Average EMG and audio signals for subject NSM. (n = 89 for the stops, 120 for the fricatives, and 28 for the affricates.)

It is useful to consider the timing relationships in somewhat greater detail. For example, we consider the timing demonstrated in the top panel of Figure 1, which shows the data for stops for subject TB. The audio envelope curves indicate that timing was similar for the voiced and voiceless consonants, since the curves follow about the same course except during the period associated with aspiration for the voiceless consonants. The EMG levels are the same during the nucleus of the preceding vowel, but the curves begin to diverge shortly before the rapid descent of the audio envelope, which indicates the beginning of the transition period. The difference between the two EMG curves reaches its maximum about when the audio envelope curves approach the baseline, the point at which stop closure has been achieved for almost all tokens. The curves then approach each other again, meeting shortly after the line-up point; they are at essentially the same level through the following vowel, diverging again as the audio envelope begins its descent for the following consonant.

For fricatives, shown in the middle panel, a similar pattern is observed, although the pattern is a little less clear because the voiceless fricatives have longer durations than their voiced counterparts. Here the EMG curves begin at the same level, during the preceding vowel. The curve for the voiceless consonant begins to increase almost immediately, however, shortly before the corresponding audio envelope begins its rapid descent. The maximum difference occurs when the audio envelope for the voiceless fricative nears its minimum, and the curves converge again, this time somewhat before the lineup point rather than after it as in the upper panel. The curves diverge again from each other in the right half of the figure, at about the same place within the cycle for the voiceless consonants (i.e., at about the lead before the rapid descent of the audio envelope).

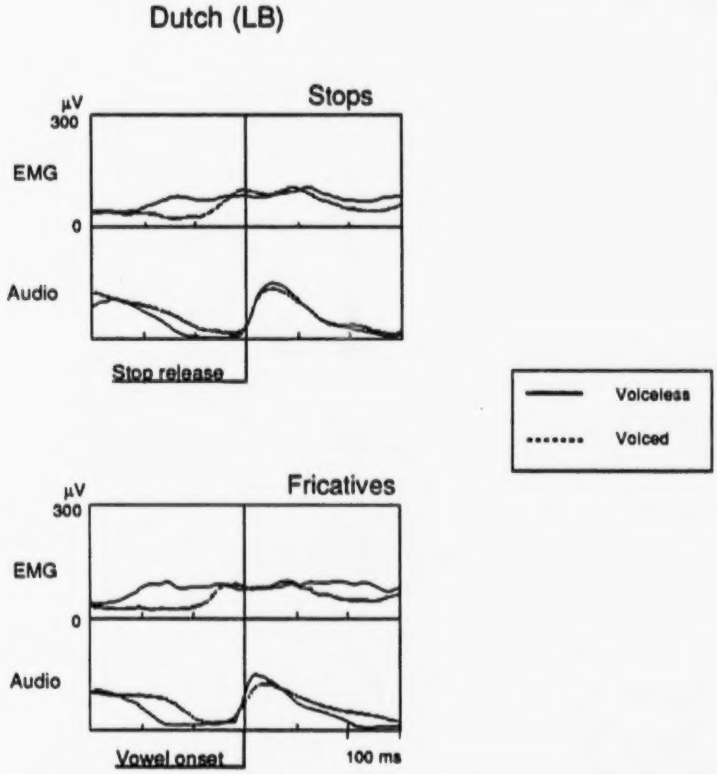


Figure 3. Average EMG and audio signals for subject LB. (n = 45 for the stops and 29 for the fricatives.)

All the curves for the other two subjects, shown in Figures 2 and 3, exhibit qualitatively the same type of behavior as Figure 1. The EMG curves diverge roughly where the audio envelope for the voiceless consonant begins its rapid descent. The maximum difference between the two EMG curves occurs around occlusion for the voiceless consonant (although not so closely for the fricatives and affricates for subject NSM in Figure 2), and the curves converge again in the vicinity of the lineup point. Subject NSM also differs somewhat from the other subjects in the right halves of the plots, in that the curves for voiced and voiceless consonants separate much earlier, during the vowel in the preceding (stressed) syllable. The results for subject LB in Figure 3 show the same timing of CT activity as those for subject TB in Figure 1.

The difference in the EMG curves was tested quantitatively using the integrated "area under the curve" measure, described in Methods. Figure 4 shows the integrated value of cricothyroid activity averaged over all tokens for each condition. Standard deviations are also indicated. In all cases except one, the affricates for subject TB, the difference between the voiced and voiceless consonants is highly significant ($p < 0.001$). For the affricates of subject TB, the difference was not statistically significant ($t_{29} = 0.43$, $p > 0.5$).

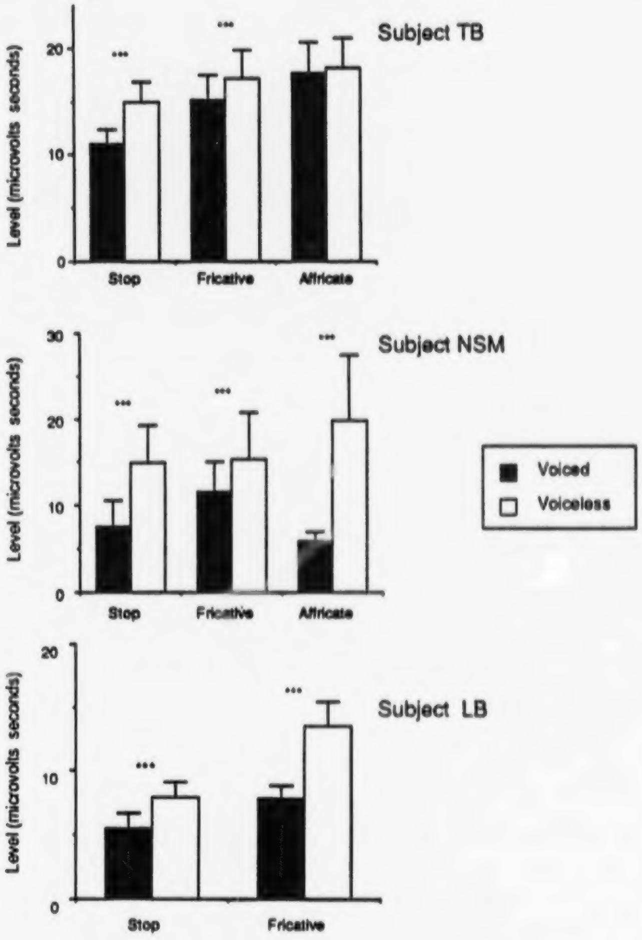
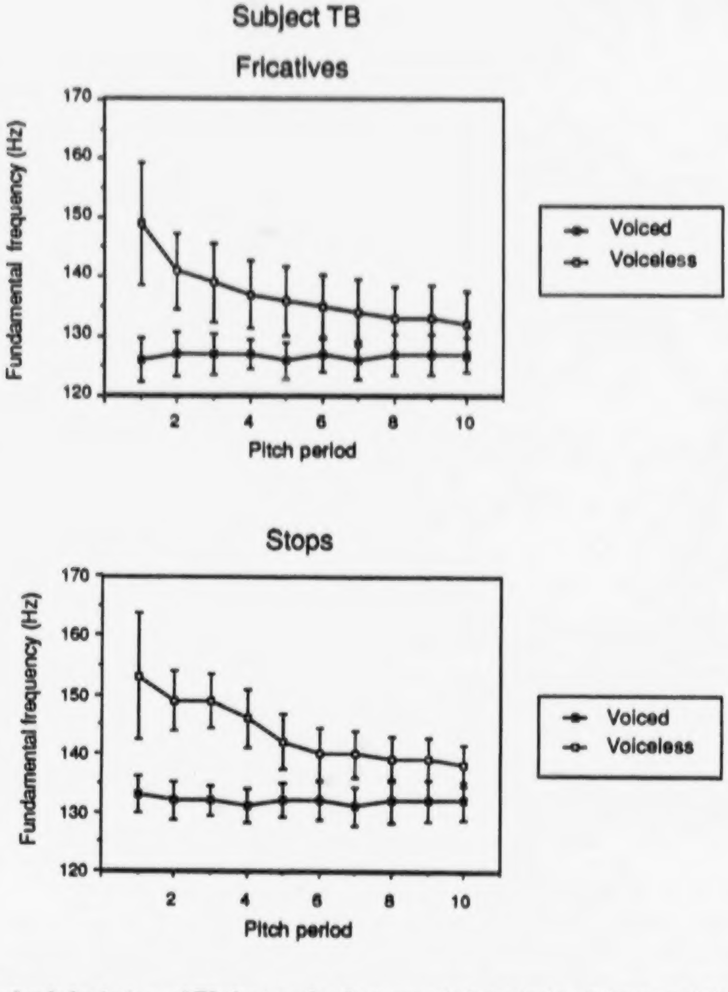


Figure 4. Mean and standard deviation of cricothyroid activity level. *** = $p < 0.001$.

The audio envelopes in Figures 1-3 suggest that the vocal folds were vibrating for the voiced consonants produced by subjects NSM and LB. On the other hand, the voiced stops for subject TB in Figure 1 would seem to have been produced at least partly devoiced. This can be deduced from the amplitude of the audio envelope during stop closure, which is almost similar for the voiced and voiceless cognates. The vibrations seem to stop during the later part of the oral closure for the voiced stops.

B. Fundamental Frequency

The measurements of fundamental frequency at the beginning of the vowel following the consonant are shown in Figures 5-7 for the stops and fricatives for the three subjects and in Figure 8 for the affricates. As expected, the F0 at this point of the vowel is higher after a voiceless consonant. In this voiceless context, the fundamental frequency generally starts at a high level and drops during the first periods. In the voiced context, F0 at the onset of the vowel is lower and may remain at the same level, rise, or even fall slightly. The difference in F0 following voiced and voiceless consonants was in most cases highly significant for all the ten periods measured. The only exception to this general trend consisted of the affricates for subject TB in Figure 8. Here, the difference disappeared at pitch period number 7 ($t_{28} = 1.26$, $p > 0.1$) and also for the preceding pitch periods the difference was less robust. Recall that for this speaker the difference in cricothyroid activity between the voiced and voiceless affricates was not significant, cf. Figures 1 and 4.


Subject TB

Fricatives

Stops

Figure 5. Mean and standard deviation of F0 during the first ten pitch periods of the vowel following stops and fricatives for subject TB. (n = 15.)

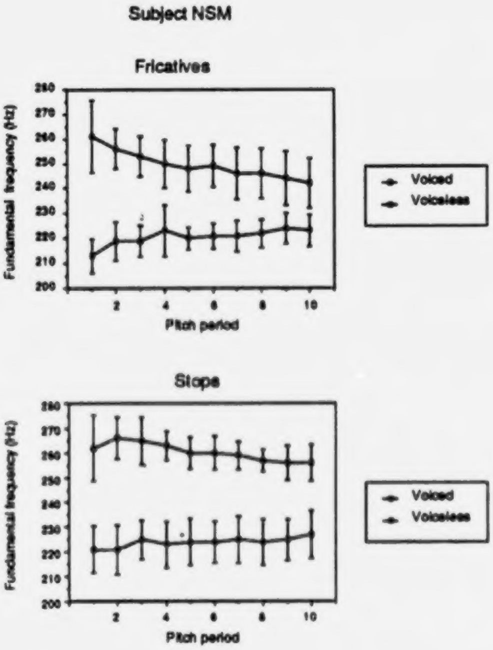


Figure 6. Mean and standard deviation of F0 during the first ten pitch periods of the vowel following stops and fricatives for subject NSM. (n = 15.)

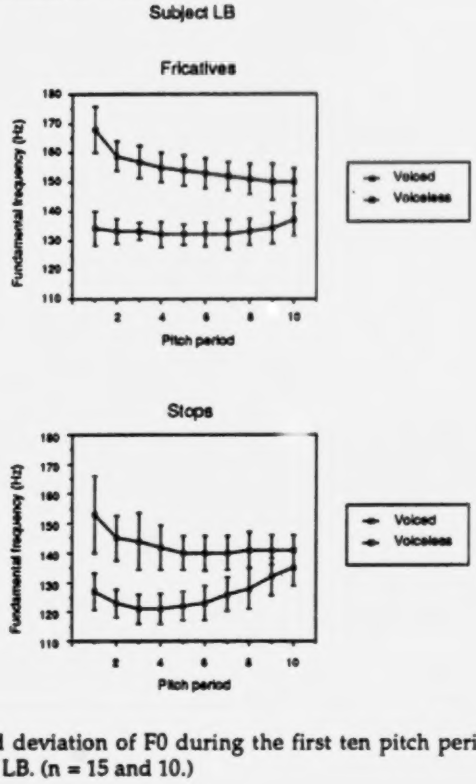


Figure 7. Mean and standard deviation of F0 during the first ten pitch periods of the vowel following stops and fricatives for subject LB. (n = 15 and 10.)

Overall, the consonant induced variations in F0 appear to be less robust following affricates, cf. Figure 8. As already noted, the difference disappeared at pitch period number 7 for subject TB. For subject NSM, the difference was highly significant for the first 6 pitch periods and significant for periods 7 through 10. It is evident, however, that for this subject, NSM, the absolute difference in Hz at the tenth period after the affricates is much smaller than at the same period following the stops and fricatives in Figure 6. Furthermore, given the higher F0 of subject NSM, her tenth pitch period occurs approximately at the same point in time after vowel onset at pitch period number 7 of subject TB.

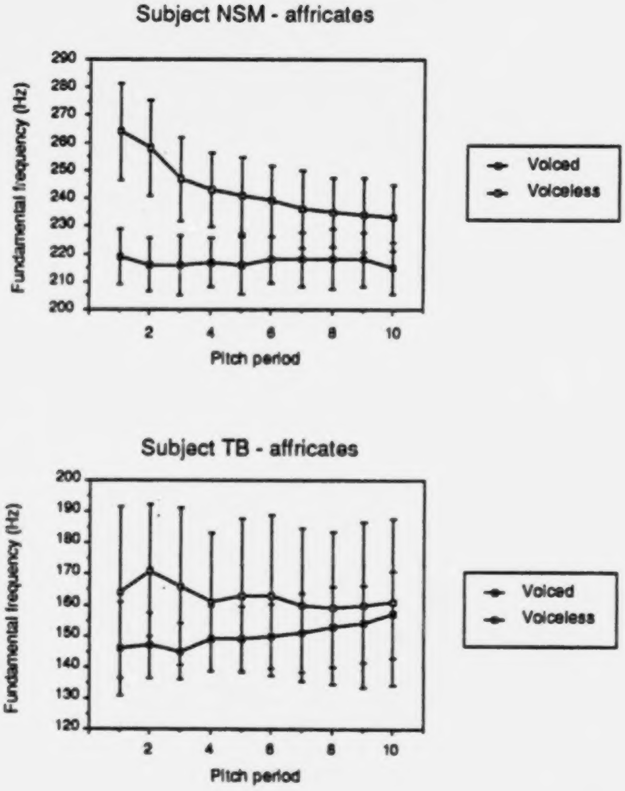


Figure 8. Mean and standard deviation of F0 during the first ten pitch periods of the vowel following affricates for subjects NSM and TB. (n = 10 and 15, respectively.)

The relationship between the activity of the cricothyroid muscle during the consonant and the fundamental frequency at the beginning of the following vowel is illustrated in Figure 9. This figure plots the average levels of the cricothyroid (cf. Figure 4) against the F0 at the onset of the following vowel; the value of F0 represents the mean of the first three pitch periods. A positive relationship is found with Pearson product moment correlation coefficients of .64, .85, and .92 for subjects TB, NSM, and LB, respectively.

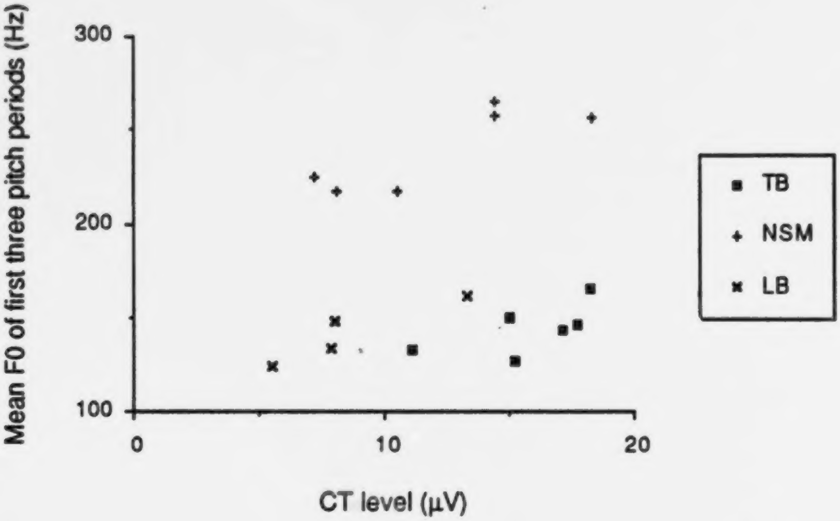


Figure 9. Plot of cricothyroid activity level versus F0 at the beginning of the following vowel.

III. DISCUSSION

The results of the present study show that the cricothyroid muscle increases its activity for voiceless consonants for all three subjects. This activity would act to increase the longitudinal tension of the cover of the vocal folds and should contribute to the inhibition of glottal vibrations. This was found for speakers of two languages with different stop systems. In both American English and Dutch, the stops are either voiced or voiceless. However, the Dutch voiceless stops are unaspirated with a VOT of about 10-15 ms in the present study while the American English voiceless stops are aspirated with a VOT of 60-80 ms (cf. Lisker & Abramson, 1964).

The present measurements of fundamental frequency at the beginning of the vowel following the voiced and voiceless consonants are in good agreement with published results (cf. Ohde, 1984; Silverman, 1986). That is, F0 is higher after a voiceless than after a voiced consonant irrespective of the manner of production, that is, stop, fricative, or affricate. This is true also for the Dutch data in Figure 7 where the voiceless stops are unaspirated; this has also been shown for French stops where the voiceless series is unaspirated (Fischer-Jörgensen, 1972). Similarly, Ohde (1984) found that the F0 of a vowel was similar if the preceding consonant was a voiceless aspirated or unaspirated stop. Thus, it appears that the conditioning factor for the F0 difference is voicing and not aspiration. This conclusion is further strengthened by the fact that studies of F0 after voiceless aspirated and unaspirated stops have shown conflicting results where interspeaker and interlanguage differences seem to occur to a great extent (cf. Gandour, 1974; Hombert & Ladefoged, 1977; Jeel, 1975; Kagaya, 1974; Kagaya & Hirose, 1975; Zee, 1980). Presumably, aerodynamic factors are also involved. However, the results presented by Hutters (1984, 1985) suggest that the cricothyroid has a higher level of activity in Danish aspirated than unaspirated stops.

If there is thus a distinct difference in the level of F0 after voiced and voiceless consonants, the pattern of pitch change at the beginning of the vowel may differ. After voiceless consonants, F0 generally drops, although this is not clear for the vowels following stop consonants for subject NSM, cf. Figure 6, or for those following affricates for subject TB, cf. Figure 8. After voiced consonants, the fundamental frequency may either rise, fall, or stay level; all these three patterns can be found in Figures 5-8. The F0 change at the beginning of the vowel thus depends not only on the preceding consonant but also on the overall intonation pattern of the utterance (cf. Silverman, 1986).

We thus argue that an increased longitudinal tension of the vocal folds is associated with voicelessness. The difference in tension between voiced and voiceless consonants is still manifest at the beginning of a following vowel thus accounting for the commonly found consonant-induced variations in F0, cf. Figure 9 that shows a positive correlation between the cricothyroid activity during the consonant and the F0 level at the beginning of the following vowel. According to the present results, there is no clear difference in cricothyroid activity during the vowel following the consonant. If variations in cricothyroid activity are associated with the voicing distinction, they should occur at the transition between the vowel and the consonant, given the fact that the electrophysiological events lead the resulting mechanical effects with the latency depending on the contraction and relaxation time of the muscle. If CT activity were meant only to control F0 and not to contribute to devoicing, it might be expected to be initiated later relative to stop closure and fricative constriction. Although somewhat conflicting, studies by Baer (1981) and by Larson and Kempster (1985) of single motor units in the human cricothyroid muscle and by Atkinson (1978) of the EMG interference pattern during utterances with various intonation patterns indicate that the effective contraction time (actually the latency between EMG and the onset of pitch increase) is in the range of 20-80 ms, and relaxation times are somewhat longer than contraction times. The present results would also rationalize why the voicing status of a consonant appears to affect the fundamental frequency of a preceding vowel, independently from the prosodic structure of the utterance, to a lesser extent than it affects the following vowel.

While we would thus suggest that the well-attested F0 difference between vowels following voiced and unvoiced consonants is due to a change in vocal fold tension caused by the cricothyroid muscle, it is obvious that other factors may also be involved. As reviewed by Hombert et al. (1979) these factors may include aerodynamic effects during stop closure and release and changes in the vertical position of the larynx that can affect the tension of vocal-fold tissues. While we have concentrated on the cricothyroid muscle, our reading of the literature suggests that the other adductor muscles of the larynx are suppressed for voiceless consonants; thus the body of the folds would be relaxed as the activity of the vocalis muscle decreases for voiceless consonants.

Kingston (1986) has suggested an alternate explanation for the consonant-voicing pitch effect. Analysis in different languages suggests that the pitch effect consistently accompanies the phonological voicing distinction, regardless of whether or not the vocal folds are actually vibrating. He suggests, then, that the pitch in the neighborhood of obstruents may be controlled independently as a cue to phonological voicing, rather than being a by-product of the gestures that control vibrations of the vocal folds and presence of aspiration. Although the present experiment did not specifically address this hypothesis, we would argue that the timing of increased CT activity for the voiceless consonants suggests that it is related to devoicing rather than to pitch control; if CT activity was only related to F0 control, the increase in activity should occur closer to the onset of the vowel following the consonant. Furthermore, the perceptual

44

effectiveness of these F0 variations in cueing consonant voicing appears to be debatable (cf. Abramson & Lisker, 1985; Silverman, 1986).

Perhaps the most compelling evidence in favor of the line of reasoning that we have presented here is the finding for the affricates of subject TB. In this case, there is no significant difference in cricothyroid activity between the voiced and unvoiced cognates. Nor is the F0 difference in the following vowel clear-cut and robust in the context of the affricates. Taken together, these results, together with a review of the literature, suggest that control of laryngeal tension by means of the cricothyroid muscle is a mechanism that may be used to supplement abduction and reduction of transglottal flow as devoicing mechanisms. Though it is evidently not an essential component of the mechanism, there is converging evidence that it is used frequently, at least in prestressed position. The EMG curves suggest that the CT is also used to contribute to consonant devoicing in unstressed syllables, but further analyses are required to determine whether the timing relationships are similar to those for stressed syllables.

ACKNOWLEDGMENT

This work was supported by NINCDS Grants NS-13617 and NS-13870 and NIH Biomedical Research Support Grant RR-05596 to Haskins Laboratories.

REFERENCES

Abramson, A., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.), *Linguistic phonetics*. New York: Academic.

Atkinson, J. (1978). Correlation analysis of the physiological features controlling fundamental voice frequency. *Journal of the Acoustical Society of America, 63*, 211-222.

Collier, R., Lisker, L., Hirose, H., & Ushijima, T. (1979). Voicing in intervocalic stops and fricatives in Dutch. *Journal of Phonetics 7*, 357-373.

Baer, T. (1981). Investigation of the phonatory mechanism. *Proceedings of the Conference on the Assessment of Vocal Pathology. ASHA Reports, 11*, 38-46.

Bell-Berti, F. (1975). Control of cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America, 57*, 456-464.

Dixit, P., & McNeilage, P. (1980). Cricothyroid activity and control of voicing in Hindi stops and affricates. *Phonetica, 37*, 397-406.

Fischer-Jörgensen, E. (1972). *PTK* et *BDG* français en position intervocalique accentuée. In A. Valdman (Ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre*, (pp. 143-200). The Hague: Mouton.

Fre Woldu, K. (1985). The perception and production of Tigrinya stops. *Reports from Uppsala University*, (Department of Linguistics), *13*, 1-190.

Gandour, J. (1974). Consonant type and tone in Siamese. *Journal of Phonetics, 2*, 337-350.

Halle, M., & Stevens, K. (1971). A note on laryngeal features. *Quarterly Progress Report, Research Laboratory of Electronics*, (MIT), *101*, 198-213.

Hirose, H., & Gay, T. (1973). Laryngeal control in vocal attack. An electromyographic study. *Folia Phoniatrica, 25*, 203-213.

Hirose, H., & Niimi, S. (1987). The relationship between glottal opening and the transglottal pressure differences during consonant production. In T. Baer, C. Sasaki, & K. S. Harris, (Eds.), *Laryngeal function in phonation and respiration* (pp. 381-390). Boston: College-Hill.

Hirose, H., & Ushijima, T. (1978). Laryngeal control for voicing distinction in Japanese consonant production. *Phonetica, 35*, 1-10.

Hombert, J-M., & Ladefoged, P. (1977). The effect of aspiration on the fundamental frequency of the following vowel. *UCLA Working Papers in Phonetics, 36*, 33-40.

Hombert, J-M., Ohala, J., & Ewan, W. (1979). Phonetic explanations for the development of tones. *Language, 55*, 37-58.

House, A., & Fairbanks, G. (1953). The influence of consonantal environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25,* 105-113.

Hutters, B. (1984). Vocal fold adjustment in Danish voiceless obstruent production. *Annual Report,* (Institute of Phonetics, University of Copenhagen), *18,* 293-385.

Hutters, B. (1985). Vocal fold adjustment in aspirated and unaspirated stops in Danish. *Phonetica, 42,* 1-24.

Jeel, V. (1975). An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. *Annual Report,* (Institute of Phonetics, University of Copenhagen), *9,* 191-211.

Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics,* 2, 161-180.

Kagaya, R., & Hirose, H. (1975). Fiberoptic, electromyographic, and acoustic analyses of Hindi stop consonants. *Annual Bulletin,* (Research Institute of Logopedics and Phoniatrics, University of Tokyo), *9,* 27-46.

Kingston, J. (1986). Are F0 differences after stops accidental or deliberate? *Journal of the Acoustical Society of America, 79,* S27 (A).

Larson, C., & Kempster, G. (1985). Voice fundamental frequency changes following discharge of laryngeal motor units. In I. Titze, & R. Scherer (Eds.), *Vocal fold physiology: Biomechanics, acoustics and phonatory control* (pp. 91-103). Denver: The Denver Center for the Performing Arts.

Lehiste, I., & Peterson, G. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America, 33,* 419-425.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20,* 384-422.

Löfqvist, A. (1980). Interarticulator programming in stop production. *Journal of Phonetics, 8,* 475-490.

Löfqvist, A., & McGarr, N. S. (1987). Laryngeal dynamics in voiceless consonant production. In T. Baer, C. Sasaki, & K. S. Harris (Eds.), *Laryngeal function in phonation and respiration* (pp. 391-402). Boston: College-Hill.

Löfqvist, A., McGarr, N. S, & Honda, K. (1984). Laryngeal muscles and articulatory control. *Journal of the Acoustical Society of America, 76,* 951-954.

Löfqvist, A., & Yoshioka, H. (1980). Interarticulator programming in obstruent production. *Phonetica, 38,* 21-34.

Löfqvist, A., & Yoshioka, H. (1984). Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. *Speech Communication, 3,* 279-289.

Minifie, F., Abbs, J., Tarlow, A., & Kwaterski, M. (1974). EMG activity within the pharynx during speech production. *Journal of Speech and Hearing Research, 17,* 497-504.

Ohde, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America, 75,* 224-230.

Perkell, J. (1976). Responses to unexpected suddenly induced change in the state of the vocal tract. *Quarterly Progress Report, Research Laboratory of Electronics, MIT, 117,* 273-281.

Sawashima, M., & Hirose, H. (1983). Laryngeal gestures in speech production. In P. MacNeilage (Ed.), *The production of speech,* (pp. 11-38). New York: Springer.

Silverman, K. (1986). FO segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica, 43,* 76-91.

Sonesson, B. (1982). Vocal fold kinesiology. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), *Speech motor control,* (pp. 113-117). Oxford: Pergamon.

Stevens, K. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica, 34,* 264-279.

Titze, I. (1980). Comments on the myoelastic-aerodynamic theory of phonation. *Journal of Speech & Hearing Research, 23,* 495-510.

Titze, I. (1986). Mean intraglottal pressure in vocal fold oscillation. *Journal of Phonetics, 14,* 359-364.

van den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research, 1,* 227-244.

Vencov, A. (1968). A mechanism for production of voiced and voiceless intervocalic consonants. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 21,* 140-144.

Westbury, J. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America, 73,* 1322-1336.

Yoshioka, H. (1981). Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica, 38,* 236-251.

Yoshioka, H. (1984). Glottal area variation and supraglottal pressure change in voicing control. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo, 18,* 45-49.

Zee, E. (1980). The effect of aspiration on the F0 of the following vowel in Cantonese. *UCLA Working Papers in Phonetics, 49,* 90-97.

FOOTNOTES

[*] *Journal of the Acoustical Society of America,* in press.

[†] Also Department of Logopedics and Phoniatrics, Lund University, Lund, Sweden.

[††] Also Center for Research in Speech and Hearing Sciences, Graduate School and University Center, The City University of New York.

[†††] Also Department of Speech and Hearing Sciences, University of Connecticut, Storrs.

Growth of Surface Waves with Application to the Mucosal Wave of the Vocal Folds

R. S. McGowan

The initiation of surface waves over a layered structure is discussed. The analysis is done on a plane layered structure consisting of flowing air, epithelium, Reinke's space, and muscle. Only after the problem is solved for the homogeneous planar structure is a length scale imposed to determine a preferred wavelength, and the conditions for instability derived. Although the layered structure is a highly idealized version of the vocal folds, this analysis illuminates one type of instability allowing for the initiation of the mucosal wave. This same linear analysis can help us understand some of the dynamics of vibration during various parts of the glottal cycle. The method used is similar to that found in the linear analysis of flutter of aircraft panels.

INTRODUCTION

The analysis performed here takes the mucosal surface wave observed during phonation to be a surface wave of a solid, elastic epithelium. The initiation of surface waves on solid or fluid surfaces over which air is flowing is analyzed in many places in the physics and engineering literature. The simplest of this type of analysis is a linear one done on layered structures when the boundaries between the layers are planar. The layers are supposed to be in some initial configuration, for example, with air at some speed flowing over an elastic surface. A time dependent disturbance is added to the operating point, the equations of motion in each layer are linearized with respect to the disturbance, and then solved with boundary conditions between the layers enforced. Because the disturbance is assumed small the linearized equations are supposed adequate to describe the physics well initially. If the disturbance is found to grow in time the system is unstable and a surface wave can be expected to result.

Rayleigh performed linearized planar analyses to study the stability of jets (Rayleigh, 1945). The layers he used were slabs of air moving at different speeds. Lamb, in his monograph on hydrodynamics, shows how linear analysis might apply to the initiation of wind waves over water (Lamb, 1945). Miles has discussed the instability in connection with wind waves over water and oil (Miles, 1959). Engineers have studied linear stability in the context of flutter of aircraft panels during flight (Miles, 1956).

48

These latter studies are the most relevant to vocal fold vibration. The analysis here closely parallels the analysis of flutter performed by Miles (1956).

The initiation of the mucosal wave is considered when the vocal folds are assumed to have a planar, layered structure (Figure 1). The layers are based on Hirano's description of the layered structure of the vocal folds. (Hirano 1981, Hirano et al., 1983). The layers going from the midline of the glottis are air, epithelium, Reinke's space, and the deep layer of ligament and muscle. The variations normal to the coronal plane are not considered. Each of these layers has bulk mechanical parameters associated with it, such as density, elasticity, temperature, and so on. It is hoped that from this idealization and with some known values of the mechanical parameters that some aspects of mucosal wave growth can be understood.

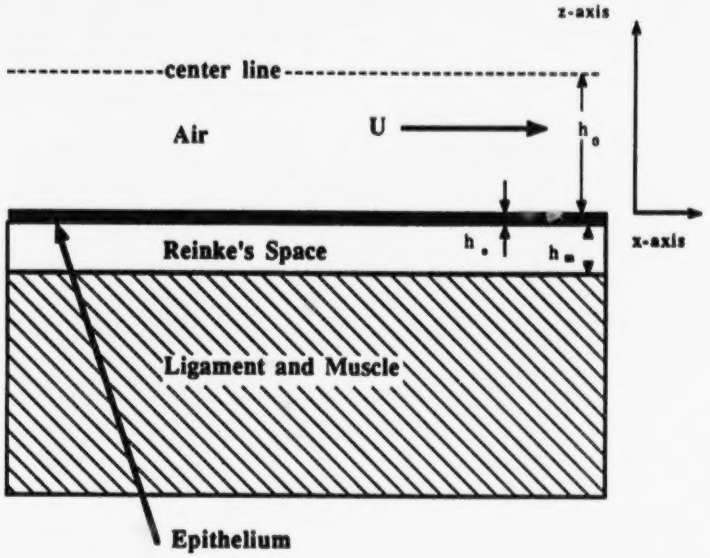


Figure 1. Layered structure of the folds. (The layers are understood to extend along the x-axis to infinity.)

The idea of using a layered structure to model the mechanical properties of the folds is similar to Titze's approach to modeling the vibration of the vocal folds (Titze, 1988; Titze & Strong, 1975; Titze & Talkin, 1979). An important difference is that a complete model along the lines of Titze's is not proposed here. This note is only intended to explore the initiation of the mucosal wave from a layered structure point of view. To this end, the analysis here takes a detailed view of the outer layers, with epithelium and Reinke's space treated as separate layers instead of then being grouped into a single mucosal layer.

Before going on to develop the equations, some limitations of the analysis should be noted. This analysis, initially, does not take account of the inhomogeneity of conditions that really do prevail in each of the layers. For instance, the speed of the air increases substantially through the glottal constriction, which means that there is a natural length scale associated with the flow of the air—the vertical length of maximum glottal constriction. Further, the layers of the folds are inhomogeneous in the vertical direction. Only the portion of the squamous epithelium covering Reinke's space will be considered capable of supporting waves. This suggests the vertical length of Reinke's space as another length scale. We will assume that the maximum glottal constriction length is shorter than the length of Reinke's space. It will be seen that the greater the air flow and the longer the surface wavelength, the more likely that instability results. (The shorter wavelength disturbances are more stable because the bending stiffness increases as the fourth power of wavenumber.) As a result, a surface wave is most likely to appear in the glottal constriction with a fundamental wavelength equal to twice the constriction length. (The bulge created in the epithelium will have the glottal constriction length as its length scale, so that the fundamental wavelength of this bulge will be twice the glottal constriction length.) So, while the analysis is initially done on an infinite layered structure, a length scale is imposed later to choose a wavelength to calculate stability regimes. Although not completely consistent, this type of analysis is common and does give us insight to the physical mechanisms responsible for surface wave initiation. Also, the viscous shear stress is neglected in this analysis.

The analysis performed here is a linear analysis, where the equations of motion have been linearized about some assumed initial configuration. If we are willing to make the quasi-steady approximation, the same analysis can be used to describe the dynamics of the vocal fold vibration, locally, in different parts of the glottal cycle.

ANALYSIS

A linear stability analysis is applied to the idealized model about an operating point. The operating point to be considered includes uniform airflow at speed U, much less than the speed of sound, a_0. The density of air is ρ_a and the viscosity of the air is neglected. The epithelium is supposed to be under tension T per unit breadth (dimension into the paper in Figure 1), and have bending stiffness, D, per unit breadth. The epithelium is presumed thin compared to any disturbance wavelength, and it has mass per unit area, m. Reinke's space consists of an incompressible fluid of density, ρ, and viscosity is neglected here. The muscle is immovable. This later condition can be made a little more realistic with a general impedance boundary condition. Given this operating point, the linear equations of motion for a small disturbance can be written.

With the neglect of viscosity, the velocity field of a small disturbance in the air can be described with a velocity potential, ϕ_a, that satisfies the convected wave equation,

$$\left(\frac{1}{a_0^2}\frac{D^2}{Dt^2} - \nabla^2\right)\phi_a = 0, \tag{1}$$

where

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + U\frac{\partial}{\partial x}.$$

If ζ denotes the inward deflection of the epithelium from rest position $z = 0$, (position when there is no pressure difference across the epithelium), then the equation of motion is,

$$m\frac{\partial^2\zeta}{\partial t^2} + D\frac{\partial^4\zeta}{\partial x^4} - T\frac{\partial^2\zeta}{\partial x^2} = -\Delta p, \tag{2}$$

where

$\Delta p = p_a|_{z=0} - p|_{z=0}$,
p_a = pressure of the air,
p = pressure in Reinke's space.

Because the fluid of Reinke's space is assumed to be inviscid and incompressible, small disturbances in this layer can be described by a velocity potential satisfying,

$$\nabla^2\phi = 0 \tag{3}$$

The boundary conditions help to determine how the disturbances in each layer interact with one another. Because there is no flow across the midline of the channel, $z = h_0$,

$$\left.\frac{\partial\phi_a}{\partial z}\right|_{z=h_0} = 0 \tag{4}$$

Similarly, the muscle is assumed to be immovable, so,

$$\left.\frac{\partial\phi}{\partial z}\right|_{z=-h_m} = 0 \tag{5}$$

At the epithelium, the velocity at the surface is continuous,

$$\frac{D}{Dt}\zeta = \left.\frac{\partial\phi_a}{\partial z}\right|_{z=0} \tag{6}$$

and

$$\frac{\partial\zeta}{\partial t} = \left.\frac{\partial\phi}{\partial z}\right|_{z=0} \tag{7}$$

The smallness of the disturbance in comparison with the wavelength of the disturbance was used to linearize these boundary conditions, at the plane of equilibrium $z = 0$. (A Taylor series expansion is used to write the boundary conditions on the physical surface at the $z = 0$ plane. Linearization allows all but the first term of the series to be neglected.)

Finally, the relationships between the pressures and velocity potentials are given by linearized Bernoulli's relations,

$$p_a = -\rho_a\frac{\partial\phi_a}{\partial t} - \rho_a U\frac{\partial\phi_a}{\partial x} = -\rho_a\frac{D}{Dt}\phi_a \tag{8}$$

and

$$p = -\rho\frac{\partial\phi}{\partial t} \tag{9}$$

The set of equations (1) through (9) forms a closed set.

To find a solution to the above set of equations, assume,

$$\zeta = \zeta_0 \exp\left(ik(ct - x)\right) \tag{10}$$

where k is the wavenumber and c is the phase speed.

Because of the boundary conditions equation (6) and equation (7), the velocity potentials must be of the form,

$$\phi_a = \tilde{\phi}_a(z) \exp\left(\iota k(ct - x)\right) \tag{11}$$

and

$$\phi = \tilde{\phi}(z) \exp\left(\iota k(ct - x)\right) \tag{12}$$

Substituting equation (11) into equation (1) gives,

$$\frac{d^2\tilde{\phi}_a}{dz^2} - k^2\left[1 - \frac{(c - U)^2}{a_0^2}\right]\tilde{\phi}_a = 0 \tag{13}$$

The solution is,

$$\tilde{\phi}_a = A'\, e^{\beta kz} + B' e^{-\beta kz}\ , \tag{14}$$

where

$$\beta^2 = 1 - \frac{(c - U)^2}{a_0^2}$$

and the constants A' and B' are determined by the boundary conditions. Applying equation (4) to the solution equation (14) gives,

$$\tilde{\phi}_a = A\left\{e^{-2\beta kh_0} e^{\beta kz} + e^{-\beta kz}\right\} \tag{15}$$

In a similar way, substituting equation (12) into equation (3), and using the boundary condition expressed in equation (5) gives,

$$\tilde{\phi} = B\left\{e^{2kh_m} e^{kz} + e^{-kz}\right\} \tag{16}$$

Combining equation (6) with equation (8) and (7) with equation (9) gives,

$$\left.\frac{\partial p_a}{\partial z}\right|_{z=0} = -\rho_a \frac{D^2\zeta}{Dt^2} \tag{17}$$

and

$$\left.\frac{\partial p}{\partial z}\right|_{z=0} = -\rho \frac{\partial^2\zeta}{\partial t^2} \tag{18}$$

Impedances at the epithelium can be found by using equations (8), (9), (15), and (16).

$$\frac{p_a|_{z=0}}{\left.\frac{\partial p_a}{\partial z}\right|_{z=0}} = \frac{-1}{\beta k \tanh\left(\beta k h_0\right)} \tag{19}$$

and

$$\frac{p|_{z=0}}{\left.\frac{\partial p}{\partial z}\right|_{z=0}} = \frac{1}{k \tanh(k h_m)} \tag{20}$$

Using equations (17) through (20) a homogeneous differential equation replacing the inhomogeneous equation (2) can be written,

$$\frac{\rho_a}{\beta k \tanh\left(\beta k h_0\right)}\frac{D^2\zeta}{Dt^2} + \left(m + \frac{\rho}{k \tanh(k h_m)}\right)\frac{\partial^2\zeta}{\partial t^2} + D\frac{\partial^4\zeta}{\partial x^4} - T\frac{\partial^2\zeta}{\partial x^2} = 0 \tag{21}$$

Substituting equation (10) into equation (21) gives,

$$\left[1 + \frac{\rho}{km \tanh(k h_m)}\right]c^2 + \frac{\rho_a}{\beta k m \tanh\left(\beta k h_0\right)}(c - U)^2 - c_0^2 = 0 \tag{22}$$

where

$$c_0^2 = \left(Dk^2 + T\right)\Big/_{m} \ .$$

Let $\mu = 1 + (\rho/km)\tanh(kh_m)$ and $\mu_a = (\rho_a/\beta km)\tanh(\beta k h_0)$. μ is, essentially, the ratio of the inertial coefficient of the epithelium plus the inertial coefficient of Reinke's space fluid to the inertial coefficient of the epithelium. μ_a is, essentially, the ratio of the inertial coefficient of the air to that of the epithelium. Equation (22) can be written,

$$(\mu + \mu_a)c^2 - (2\mu_a U)c + \left(\mu_a U^2 - c_0^2\right) = 0 \tag{23}$$

In the following it is assumed that U/a_0, $c/a_0 << 1$. The air is considered to be incompressible, both because the Mach number of the mean flow is small, and because the length scale of the glottal constriction is small compared to any acoustic wavelength under consideration. (The length scale of concern is the length of the glottal constriction because this is where the maximum velocity occurs. This will be shown to be the most likely place for an instability.) The assumption of incompressibility means that β is approximately equal to one, and that a simple closed form solution for the surface wave speed can be obtained. Otherwise, only approximate solutions are available.

Solving for the phase speed,

$$c = \frac{\mu_a}{\mu + \mu_a}U +- \frac{\sqrt{c_0^2(\mu + \mu_a) - \mu\mu_a U^2}}{\mu + \mu_a} \tag{24}$$

c_0 is the natural phase speed of surface waves on the epithelium without the interaction of the air and Reinke's space. The larger c_0 the greater the tension of the epithelium. The effect of airflow is to decrease the ambient pressure (the Bernoulli effect), and hence counter the tension of the epithelium. As the speed of the air increases, the discriminant in equation (24) becomes negative, the phase speed becomes complex, and there is exponential growth in time.

At the point where the discriminant goes to zero, the critical point, the stiffness of the epithelium is just large enough to balance the inward force provided by the air. The air speed at the critical point is given by,

$$U_{crit} = \sqrt{\frac{1}{\mu} + \frac{1}{\mu_a}}\, c_0 \tag{25}$$

At the critical point the surface wave phase speed is given,

$$c_{crit} = \frac{1}{1 + \dfrac{\mu}{\mu_a}} U_{crit} \tag{26}$$

The real part of the phase speed can be sketched as a function of fluid speed, U (Figure 2). Below the critical fluid speed there are two real values for the phase speed. These have opposite signs for U small enough, so that there is both a stable upstream and stable downstream surface wave. However, just below the critical point both stable waves are moving downstream, and at the critical point there is only one phase speed.

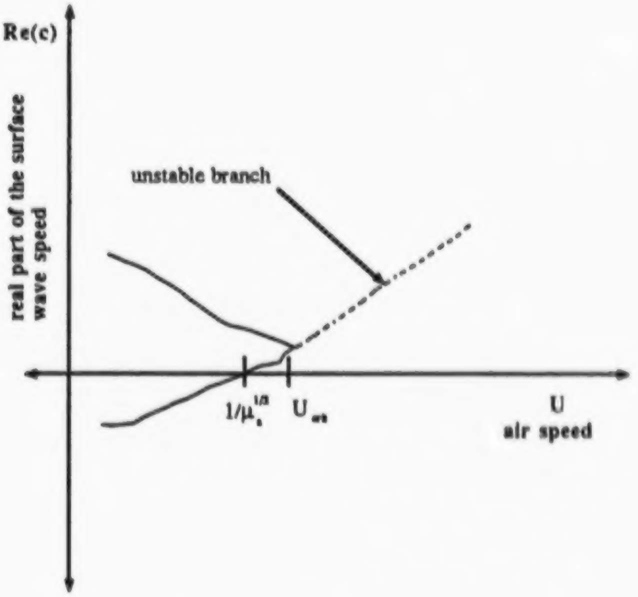


Figure 2. Stability Diagram.

It is interesting to note the dependence of the critical point on the geometric parameters and on the wavelength. These quantities appear in the inertial coefficients μ and μ_a. As μ and μ_a increase the critical air speed decreases (see equation 25). Also, μ and μ_a are increasing functions of wavelength, so the longer wavelengths require less air speed for instability. Further, as the air channel gets narrower, that is as kh_0 gets smaller, the critical air speed becomes smaller. A similar effect occurs to a much smaller magnitude with the thinning of Reinke's space. (The magnitude of the effect is much smaller because μ is so much larger than μ_a, as is shown below.)

Based on some estimates of the parameters appearing in equations (25) and (26), the critical speeds can be estimated. The air speed is large only through the glottal constriction, and because instabilities depend on the air speed being above a particular critical value, the glottal constriction length is the length scale of interest. Further, the longer wavelengths are the most likely the ones to be excited, so that a wavelength twice the glottal constriction length is a reasonable estimate of surface wavelength. The rest of the parameters are set as follows:

vertical length of glottal constriction = 1 cm
$\rho = 1\ g/cm^3$
$\rho_a = 10^{-3}\ g/cm^3$
density of epithelium = $1\ g/cm^3$
thickness of epithelium = 5×10^{-3} cm (Titze & Durham, 1987)
this implies $m = 5 \times 10^{-3}\ g/cm^2$
$h_m = 5 \times 10^{-2}$ cm (Hirano, et al., 1983)
$h_0 = 10^{-1}$ cm
$T = 5 \times 10^3\ g/sec^2$ (Titze & Durham, 1987)

The bending stiffness of the epithelium is neglected and the tension is taken to be that measured at 20% strain.

With these parameter settings equation (25) gives $U_{crit} = 2000.$ cm/sec and equation (26) gives $c_{crit} = 1.$ cm/sec. The critical air speed shown above can be attained in the glottis indicating that the type of instability derived here can lend to the initiation of the mucosal wave. However, the surface wave speed at the critical point shown above is a couple of orders of magnitude less than that observed by Baer (1975) (The surface wave speed reported by Baer was about 1 ms.) This indicates that other mechanisms are also contributing to surface wave growth.

DISCUSSION

It is interesting to compare this layered structure approach to the familiar two-mass model. The analysis of the two-mass model shows what needs to be included into this layered structure approach to make it a complete model of mucosal wave instability. Also, it will be seen that the analysis given here can provide insight into the mucosal wave itself and the dynamics of the two-mass model.

The layered structure approach is a continuum mechanics approach, while the two-mass model takes a lumped element approach. The latter has the advantage of easily incorporating the changing properties along the vocal-tract axis, although in a discontinuous way. With the layered structure approach, it is less natural to include these inhomogeneities, but it can be done. The premier advantage of the layered structure approach is the straightforward correspondence between the parameters of the actual folds and the structures in the layers.

It has been shown that mechanical loads must be included in any model of vocal fold vibration to account for the energy exchange between the air and the folds (Stevens, 1977). These loads are necessary to provide for differences between the closing movement and opening movement of the folds in the relation between pressure and surface position. This accounts for the net energy input to the folds from the air. The mechanical loads can be inertial loads of the vocal-tract air or the nonlinear resistances resulting in pressure head losses at the entrance and exit of the glottal regions (Titze, 1988). The two-mass model incorporates both loading effects. A layered structure approach can incorporate both effects also (Titze & Talkin, 1979). Therefore, a layered structure approach in the study of mucosal wave initiation can incorporate these loading effects.

After filling in the noted deficiencies, the resulting layered structure model would be a more complete linear picture of the initiation of the mucosal wave. In fact, the linear instability derived above is not necessary for the initiation of the mucosal wave. All that is needed for linear instability is that there be a negative damping coefficient, but positive restoring force during some part of the glottal cycle. (This instability is analogous to what may be called a dynamic instability in a mass-spring system with negative damping and positive spring constant (Miles, 1959).) In this paper a more catastrophic type of instability has been derived, where the effective restoring force of the epithelium and air system is negative. (This is analogous to what may be called a static instability in a mass-spring system with negative spring constant (Miles, 1959).) In the first case of the dynamic instability, exponentially growing oscillation are the result, and in the latter case of static instability, exponential growth without oscillation results. Titze has shown how the dynamic instability works on a layered structure picture of the vocal folds without "...the underlying mechanisms of surface-wave propagation." (Titze, 1988, p 1540). These mechanisms have been considered here, but with the more severe static instability caused by the Bernoulli pressure overcoming the epithelium tension.

The static instability has not been considered to be important in the analysis of models in the past. Ishizaka & Matsudaira (1972) could have found an instability with an effective negative spring force on the lower mass in their analysis of the two-mass model if they had used a slightly greater subglottal pressure, slightly more compliant springs, and a smaller rest area then the one they used. Titze (1988) recognizes the possibility of such an instability, but removes its consideration from linear problems. It has been shown here that static instabilities can be considered in a linear analysis, where, in fluid mechanics, it is known as a Kelvin-Helmholtz instability (Chandrasekhar, 1981).

This linear analysis can be applied to the parts of the glottal cycle making the quasi-steady approximation, and linearizing the equations of motion about the operating point of instantaneous glottal opening and air velocity. The static instability is most likely to appear in such an analysis during the closing phase of the glottal cycle. The severity and duration of the static of instability may be used to distinguish some voice types. For instance, the conditions of relatively loose epithelium and small (but positive) rest area may describe pressed voice. One would expect that the static instability is more severe and takes a greater portion of the closing phase in the pressed voice than in a chest voice. This offers an alternative explanation to Stevens' account of some of the aspects of pressed phonation, where he assumes that the rest area is negative for pressed voice to obtain the abrupt closure (Stevens, 1988).

To conclude, this layered structure approach may find application in other areas of communicative sound production such as syringeal vibration and trills of various articulators in the vocal tract.

ACKNOWLEDGMENT

The author gratefully acknowledges the support of NIH Grants HD-1994, NS-13870, and NS-13617 to Haskins Laboratories. The author also thanks Dr. Thomas Baer for insightful comments.

REFERENCES

Baer, T. (1975). Investigation of phonation using excised larynxes, Ph.D. Thesis, M.I.T.

Chandrasekhar, S. (1981). *Hydrodynamic and hydromagnetic stability* (pp. 481-92). New York: Dover.

Hirano, M. (1981). *The clinical examination of voice* (pp. 4-7). Vienna: Springer-Verlag.

Hirano, M., Matsuo K., Kakita, Y., Kawasaki, H., & Kurita, S. (1983). Vibratory structure vs. the structure of the vocal folds. In I. R. Titze & R. C. Scherer (Eds.), *Vocal fold physiology,* (pp. 26-40). Denver: The Denver Center for the Performing Arts.

Ishizaka, K. & Matsudaira, M. (1972). *Fluid mechanical considerations of vocal cord vibration* (Monograph 8). Santa Barbara, CA: Speech Communications Research Laboratory, Inc.

Lamb, H. (1945). *Hydrodynamics.* (pp. 461-62). New York: Dover Publications.

Miles, J. W. (1959). On the generation of surface waves by shear flow: Part 3. Kelvin-Helmholtz instability. *Journal of Fluid Mechanics, 6,* 583-98.

Miles, J. W. (1956). On the aerodynamic instability of thin panels. *Journal of the Aeronautical Sciences, 23,* 771-80.

Rayleigh, J. W. S. (1945). *Theory of sound* (Vol. 2, pp. 376-400). New York: Dover Publications.

Stevens, K. N. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica, 34,* 264-79.

Stevens, K. N. (1988). Modes of vocal fold vibration based on a two-section model. In O. Fujimura (Ed.), *Vocal physiology: Voice production, mechanisms and functions* (pp. 357-71). New York: Raven Press.

Titze, I. R. & Strong W. J. (1975). Normal modes in vocal cord tissues. *Journal of the Acoustical Society of America, 57,* 736-44.

Titze, I. R. & Talkin, D. T. (1979). A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. *Journal of the Acoustical Society of America, 66,* 60-74.

Titze, I. R. & Durham, P. L. (1987). Passive mechanisms influencing fundamental frequency control. In T. Baer, C. Sasaki, & K. Harris (Eds.), *Laryngeal function in phonation and respiration* (pp. 304-19). Boston: College -Hill Press.

Titze, I. R. (1988). The physics of small-amplitude oscillation of the vocal folds. *Journal of the Acoustical Society of America, 83,* 1536-52.

Effects of Preceding Context on Discrimination of Voice Onset Times*

Bruno H. Repp and Hwei-Bing Lin[†]

When discriminating pairs of speech stimuli from an acoustic voice onset time (VOT) continuum (e. g., one ranging from /ba/ to /pa/), English-speaking subjects show a characteristic performance peak in the region of the phonemic category boundary. We demonstrate that this "category boundary effect" is reduced or eliminated when the stimuli are preceded by /s/. This suppression does not seem to be due to the absence of a phonological voicing contrast for stop consonants following /s/, since it is also obtained when the /s/ terminates a preceding word and (to a lesser extent) when broadband noise is substituted for the fricative noise. The suppression is stronger, however, when the noise has the acoustic properties of a syllable-initial /s/, everything else equal. We hypothesize that these properties make the noise cohere with the following speech signal, which makes it difficult for listeners to focus on the VOT differences to be discriminated.

INTRODUCTION

One of the most reliable findings in speech perception research is the so-called category boundary effect for stimuli varying in voice onset time (VOT) (Wood, 1976): Subjects find it easier to discriminate syllables that fall on opposite sides of the phonemic category boundary on a VOT continuum than stimuli that, although acoustically different to a similar degree, are drawn from within a phoneme category. Such a peak in the discrimination function along an acoustic speech continuum is one of the hallmarks of categorical perception (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). Two alternative explanations of this effect have been proposed (see reviews by Howell & Rosen, 1985, and Repp, 1984). According to one, there is a psychoacoustic discontinuity along the VOT dimension that gives rise to the discrimination peak and also determines the location of the phonemic category boundary (usually between 20-40 ms of voicing lag for English-speaking listeners). According to the other explanation, untrained listeners base their discrimination judgments less on auditory qualities than on the phonemic categories assigned to the stimuli. In this view, the discrimination peak is not due to a psychoacoustic discontinuity but to subjects' attention to a higher-level, discrete representation of the input, and the location of the category boundary is determined by language-specific factors, not universal auditory ones.

There is some evidence that attention to phonological categories plays a role in VOT discrimination tasks. For example, listeners trained to pay attention to the auditory properties of the stimuli in a low-uncertainty task tend to show a reduced category boundary effect or none at all (Carney, Widin, & Viemeister, 1977; Kewley-Port, Watson, & Foyle, 1988; Sachs & Grant, 1976; Samuel, 1977; Soli, 1983; however, see also Macmillan, Goldberg, & Braida, 1988), and speakers of languages whose stop consonant voicing contrasts differ from English may show a category boundary effect at a different VOT than is characteristic for English listeners (Williams, 1977). However, there is also a history of experimentation concerning possible psychoacoustic discontinuities on VOT continua (reviewed by Repp, 1984; Howell & Rosen, 1985; Rosen & Howell, 1987a, 1987b), which culminated in findings that nonhuman animals show enhanced sensitivity to VOT differences in the region of the English phoneme boundary (Dooling, Okanoya, & Brown, 1989; Kuhl, 1981; Kuhl & Padden, 1982). The most recent contributions to this issue stem from Kewley-Port et al. (1988) and Macmillan et al. (1988). Kewley-Port et al. found that trained human listeners in a low-uncertainty task exhibited no discrimination peak along a VOT continuum; they concluded that the peak has a phonological origin. However, Macmillan et al. sampled the continuum more finely in an otherwise similar experiment and found a clear peak at about 18 ms of VOT. Thus the category boundary effect in that region of a VOT continuum does seem to have a psychoacoustic underpinning. At the same time, it is also clear from Macmillan et al.'s work that subjects do make use of "context coding" or labeling in high-uncertainty tasks such as the frequently used ABX paradigm. Such attention to phonological categories might magnify the psychoacoustic discrimination peak (if psychoacoustic and phonological effects are additive) or substitute a peak of different origin (if psychoacoustic and phonological effects are mutually exclusive, resulting from attention to independent internal representations).

The present series of experiments began as an attempt to eliminate the contribution of phonemic labeling to VOT discrimination performance in a high-uncertainty discrimination task. The method, explained below in more detail, entailed preceding the test syllables with a fixed precursor that neutralized the phonological voicing contrast on the following stop consonant. This procedure also raised the question, however, to what extent the precursor might interfere with the auditory processing of VOT through some form of forward masking or interference in auditory memory. Several additional experiments were conducted to address this issue. It was expected that an investigation of the relative sensitivity of the VOT category boundary effect to preceding context would provide new information about its origins in phonemic labeling and/or in the auditory representation of VOT.

EXPERIMENT 1

Well-known methods for manipulating the category boundary effect in speech discrimination include the substitution of analogous nonspeech stimuli, the use of listeners with different instructions or auditory skills, and the use of discrimination tasks with varying memory demands. In Experiment 1, however, the same stimuli were presented to the same listeners in the same task with the same instructions. The critical manipulation concerned the presence or absence of immediately preceding phonetic context. In one condition, the VOT differences to be discriminated were often phonemically distinctive, whereas in the other they were not. This was achieved by first presenting stimuli from a standard [pa]-[pʰa] (phonemically, /ba/-/pa/) continuum varying in VOT, and by then preceding these stimuli with a constant [s] noise plus silence appropriate for a labial closure interval. In English there is no phonemic voicing distinction for stops after tautosyllabic /s/, and although stop consonants are produced without aspiration in these clusters, the conventional orthographic transcription—and the phoneme category assigned by linguists—is /p,t,k/, not /b,d,g/ (see Lisker, 1984). This fact was exploited previously by Sawusch and Jusczyk (1981) in a study of the auditory versus linguistic origins of selective adaptation and contrast effects along a VOT continuum. For our present purpose the transcription is relevant in so far as it preempts the "voiceless" symbols and thus impedes a categorical distinction between unaspirated and aspirated stops following /s/, at least for listeners without phonetic training. To the unsophisticated listener, both [spa] and [sp^{h}a] are /spa/.

The predictions were thus very straightforward: When asked to discriminate stimuli from the [pa]-[p^{h}a] series, subjects should exhibit the typical category boundary effect; but for the [spa]-[sp^{h}a] stimuli the discrimination peak should disappear if it is due to subjects' attention to phonemic categories.[1] If, on the other hand, the category boundary effect is partially or entirely due to a psychoacoustic discontinuity (such as a "temporal order threshold" for VOT—see Pisoni, 1977, but also Rosen & Howell, 1987b), the category boundary effect should persist. This could be either because listeners always make auditory discriminations and phonological categories are merely grafted onto the auditory representations, or because listeners' attention is drawn to auditory differences in the absence of phonological contrast.

Two methodological precautions were taken against possible complications in this simple design. First, a category boundary effect might be obtained for the [spa]-[sp^{h}a] series because subjects fail to integrate the [s] noise with the rest of the syllable. This might occur because, in the rapid successive presentation of stimulus triplets for discrimination, the [s] noises may form a separate acoustic stream (Bregman, 1978; Cole & Scott, 1973). Indeed, Diehl, Kluender, and Parker (1985) argued that such streaming occurred in the above-mentioned study by Sawusch and Jusczyk (1981), and may have invalidated some of their conclusions. Sawusch and Jusczyk had used an interstimulus interval (ISI) of 300 ms. To counteract stream formation, a relatively long (1 s) interstimulus interval was used in the present discrimination task. Since, in addition, there was a 3 s response interval after each stimulus triplet, auditory streaming was considered quite unlikely under these conditions. That listeners would be able to ignore the /s/ deliberately seemed unlikely in view of Repp's (1985) demonstration that, for untrained listeners at least, it is very difficult to segregate an initial [s] noise intentionally from following phonetic context.

Second, it is possible that an [s] noise preceding [pa]-[p^{h}a] stimuli interferes with VOT perception at a strictly auditory level, through some form of forward masking or by increasing the load on auditory memory and thereby making small stimulus differences less accessible. Although it would be surprising if such interference removed the psychoacoustic category boundary effect completely, a lowering of discrimination performance and a consequent reduction of the boundary peak might occur. To assess this possibility, a third condition was included in the experiment, in which a burst of white noise preceded the [pa]-[p^{h}a] stimuli. The white noise was chosen to be at least equal in energy to the [s] noise across the whole frequency spectrum, so that its auditory interference with VOT perception would be at least equal to that caused by the [s] noise. The subjects, however, were expected to hear these stimuli as a nonspeech noise followed by /ba/ or /pa/, so a phonological category boundary effect should be obtained, perhaps attenuated by auditory interference, which indirectly would hamper labeling accuracy. Any additional reduction in the category boundary effect in the [s] noise precursor condition relative to the white noise precursor condition may then be attributed specifically to the neutralization of the phonological contrast, and hence to subjects] attention to linguistic stimulus attributes.

Methods

Subjects

Twelve Yale undergraduates were paid to participate. They were all native speakers of American English.

Stimuli

The [pa]-[p^{h}a] CV continuum was constructed on the Haskins software synthesizer in its serial configuration. Using conventional procedures, VOT (i.e., the duration of the initial aperiodic excitation) was varied from 0 to 70 ms in 10 ms steps, resulting in 8 stimuli. The syllables did not have any release bursts. In the [spa]-[sp^{h}a], or [s]-CV, continuum the stimuli were prefixed with a 58 ms natural-speech [s] noise.[2] A 60-ms silent interval intervened between the constant [s] noise and each synthetic syllable. In the noise-CV condition, a 58-ms noise burst preceded each syllable by 60 ms. This noise burst was excerpted from broadband noise recorded from a General Radio 1390-A noise generator, and its amplitude was adjusted until its flat spectral envelope (obtained by Fourier analysis) completely subsumed the typical S-shaped envelope of the [s] noise, obtained at its most intense point. Consequently, the two noises were similar in energy around 4 kHz, but the white noise had stronger low-frequency components than the [s] noise. In addition, the white noise had an abrupt onset and offset, whereas the [s] noise was naturally tapered. These differences were thought to enhance any auditory interference due to the white noise, relative to that caused by the [s] noise. The test of the phonological hypothesis was thus rather conservative.

All stimuli were digitized at a 10 kHz sampling rate with appropriate low-pass filtering at 4. 9 kHz. In each of the three stimulus conditions, all two-step (i.e., 20-ms VOT) pairings of the stimuli were presented in an AXB format. This led to 6 (stimulus pairs) x 4 (arrangements within an AXB triad) = 24 stimulus triplets, which were recorded three times in random order on magnetic tape. The ISIs were 1 s within triplets, 3 s between triplets, and 10 s between blocks.

Procedure

The tapes were played back binaurally over TDH-39 earphones in a quiet room at a comfortable intensity. Each subject listened first to the CV series, which was preceded by four easy practice trials. The task was to listen carefully to the onsets of the syllables, and to write down "A" or "B" depending on whether the second stimulus in an AXB triplet matched the first or the third, which were known to be always different from each other. Subsequently, half the subjects listened to the [s]-CV series and then to the noise-CV series, and the other half listened in the reverse order. They were told that exactly the same differences were to be discriminated, but that all syllables would be preceded by a constant [s] or noise burst, which was to be ignored.

Results and Discussion

The results, averaged across subjects, are shown in Figure 1. As expected, the discrimination function for the CV continuum exhibited a pronounced peak, suggesting a category boundary at approximately 24 ms of VOT. By contrast, in the [s]-CV condition, there was no peak at all; the discrimination function was fairly flat and performance was poor, though above chance (except for the 50/70 stimulus pair). Finally, in the noise-CV condition, there was a peak, but it was lower and narrower than the peak in the CV condition, mainly because of a large reduction in correct responses for the 10/30 pair. Note that, with the exception of the 50/70 stimulus pair, the performance decrements were restricted to the region of the original peak, even though there was some room for decrements elsewhere.

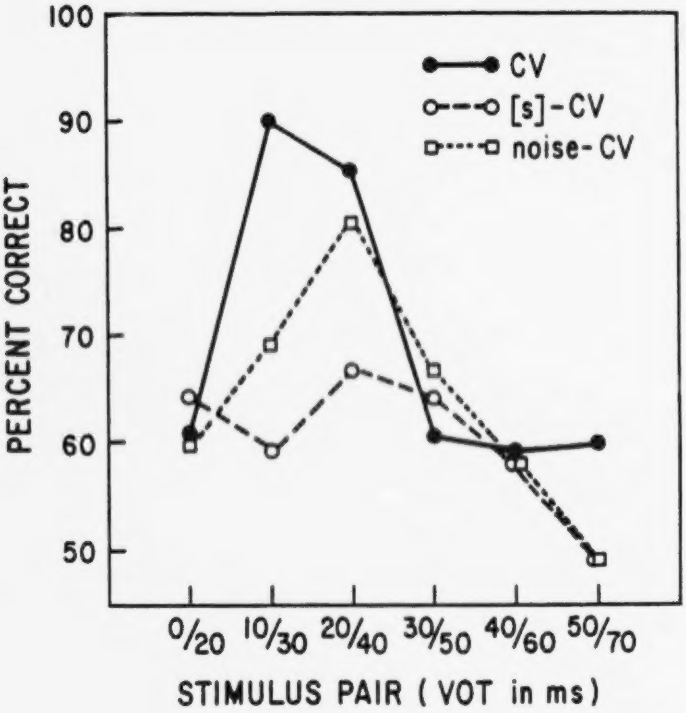


Figure 1. Discrimination performance as a function of stimulus pair in the three conditions of Experiment 1.

A repeated-measures analysis of variance was conducted on the response percentages, with the factors stimulus pair and condition. Both factors had significant main effects, $F(5,55) = 9.71$, $p < .0001$, and $F(2,22) = 6.33$, $p = .0067$, respectively, and their interaction was significant as well, $F(10,110) = 3.02$, $p = .002$. Separate analyses of pairs of conditions revealed that, although the difference between the CV and noise-CV conditions was statistically reliable, that between the noise-CV and [s]-CV conditions was not, due to considerable variability among subjects.[3]

It is clear from these results that a preceding nonspeech noise interferes with the discrimination of VOT near the category boundary. If this interference takes place in auditory processing, then the [s] noise presumably generates a similar auditory disturbance. The complete disappearance of the category boundary effect in the [s]-CV condition is consistent with the phonological hypothesis, which claims that when no phonemic contrast is perceived, there is no category boundary effect. However, the absence of a statistically reliable difference between the two precursor conditions raises the possibility that both types of noise had their effects at auditory levels of processing. Alternatively, it is possible that subjects interpreted (consciously or unconsciously) the white noise as a fricative (e. g., [ʃ]), or perceptually "restored" a fricative noise potentially hidden in the white noise (see Warren, 1984), which also would have attenuated the difference between the two precursor conditions. In that case, the results would be entirely consistent with the phonological hypothesis.

In addition to this ambiguity of interpretation, the comparison between the [s]-CV and noise-CV conditions was inherently problematic because the noises differed in a number of acoustic properties. These difficulties are endemic to studies using nonspeech substitutes for speech sounds. It was decided, therefore, to conduct a second experiment in which the difference between two precursor conditions was in the context preceding the /s/.

EXPERIMENT 2

In this experiment the syllables carrying the VOT variation were preceded by exactly the same [s] noises in both precursor conditions, but in one condition preceding context made the /s/ the initial phoneme of the test word while in the other the context made it appear to be the final phoneme of a preceding word. In the first context the phonemic voicing contrast thus was neutralized, whereas in the second it was not. The prediction of the phonological hypothesis was, therefore, that the category boundary effect should be absent in the first condition but not in the second. The auditory interference hypothesis, on the other hand, predicted similar performance in the two conditions, worse than in a control condition without preceding [s].

Methods

Subjects

Twelve new Yale undergraduates were paid to participate. All were native speakers of American English.

Stimuli

This experiment, and all following ones, used stimuli constructed entirely from natural speech. A female speaker was recorded saying the phrases *A crazy spin*, *Take this bin*, and *Take this pin*. The speech was digitized at 20 kHz with low-pass filtering at 9. 8 kHz. With the help of a waveform-editing program, the *bin* and *pin* syllables were excerpted from the *Take this* context, and a VOT continuum was fashioned by replacing initial waveform segments of *bin* with corresponding amounts of aperiodic waveform from *pin*, proceeding in steps of two glottal cycles. This resulted in stimuli with VOTs of 10, 18, 27, 36, 44, 53, and 61 ms (rounded to the nearest ms). An additional stimulus with zero VOT was created by excising the 10 ms release burst of *bin*.

The *pin* portion of *A crazy spin* was excised and discarded, leaving *A crazy s*. To make the [s] noises in the two precursors identical, the [s] of *A crazy s*, 148 ms in duration, was excerpted (without deleting it) and substituted for the shorter (94 ms) [s] noise in *Take this*. When the precursors were recombined with *bin/pin*, this was found to result in more naturally-sounding stimuli than the reverse substitution. Thus the [s] noises in both precursors had phonetic properties characteristic of syllable-initial /s/. The *Take this* precursor used derived from the *bin* context; the other *Take this* carrier phrase was discarded.

Each of the three stimulus conditions contained five blocks of 24 AXB triads presenting two-step discriminations (here corresponding to differences of about 17 ms of VOT). In the CVC condition, the *bin-pin* stimuli occurred in isolation with ISIs of 1 s within triads. In the *A crazy s*, or #[s-]CVC, and *Take this*, or [s-]#CVC, conditions, they were preceded by the respective constant precursors.[4] The silent interval between the end of the [s] noise and the CVC word was 77 ms, which equalled the original closure interval in *A crazy spin*. The ISIs within triads were 500 ms, to compensate for the longer stimulus durations. The ISIs between triads were 3 s.

Procedure

The procedure was the same as in Experiment 1, with the CVC condition first and the order of the other two conditions counterbalanced across subjects.

Results and Discussion

The results are shown in Figure 2, which is analogous to Figure 1. In the CVC condition, a pronounced peak in discrimination performance was obtained, suggesting a category boundary around 28 ms of VOT.[5] The only unusual feature of this discrimination function is the elevated performance for the first stimulus pair. This is likely to be an artifact of stimulus construction: It will be recalled that the zero VOT stimulus was generated by a different method—removal of the release burst—which gave it a "softer" onset that some subjects found very distinctive, while others did not seem to notice it. As to the #[s-]CVC and [s-]#CVC conditions, it is evident that the discrimination peak was severely reduced or absent in both, and that there was little difference between them, except perhaps for the anomalous first stimulus pair.

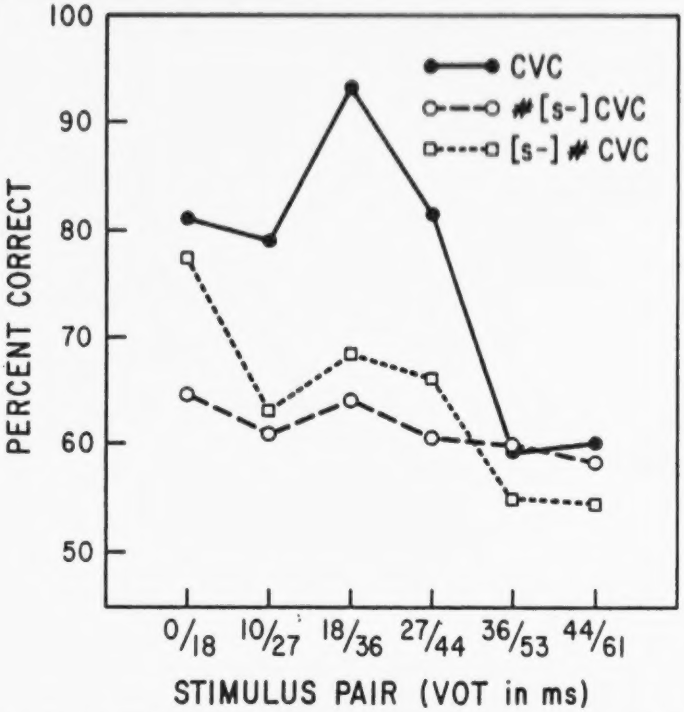


Figure 2. Discrimination performance as a function of stimulus pair in the three conditions of Experiment 2. The precursors are *A crazy s* (circles) and *Take this* (squares).

The statistical analysis revealed significant effects of stimulus pair, *F*(5,55) = 12. 40, *p* < .0001] and of condition *F*(2,22) = 34. 79, *p* < .0001, as well as a significant interaction, *F*(10,110) = 4. 62, *p* < .0001. When the CVC condition was omitted, however, only the effect of stimulus pair was significant, *F*(5,55) = 5. 02, *p* = .0007. The #[s-]CVC and [s-]#CVC results thus were statistically equivalent.

These results replicate the earlier finding that a preceding [s] noise severely reduces or even eliminates the VOT category boundary effect. However, there was no indication of any effect of linguistic structure: A preceding /s/ had the same detrimental effect, whether it initiated the test word or whether it terminated the preceding word. This seems to disconfirm the phonological hypothesis and support the auditory interference hypothesis, although the total disappearance of the boundary peak is somewhat surprising from an auditory perspective.

This interpretation of the results assumes that the phonological structure was perceived in accordance with the context but proved irrelevant to VOT discrimination. It is possible, however, that the subjects never perceived the intended difference in word boundary location and heard *Take this bin* as *Take the spin*, even though the instructions said that the precursor was *Take this*. If this seems implausible, it is still possible that VOT discrimination is performed at a prelexical level of phonological coding that depends solely on the phonetic properties of the speech segments and precedes the assignment of lexical word boundaries. The fact that aspiration noise following a stop closure is a word boundary cue (Christie, 1974) is not in contradiction with this hypothesis: Discrimination of relatively long VOTs was not affected by precursors. Thus the phonological hypothesis is still alive.

EXPERIMENT 3

To the extent that the phonetic properties of the speech segments forced a particular phonological structure on the speech signal, which took precedence over contextual constraints, Experiment 2 did not achieve its purpose. It merely confirmed the basic finding that a preceding [s] noise with syllable-initial phonetic properties eliminates the category boundary effect. Would an [s] precursor that *unambiguously* terminates a preceding word still interfere with VOT discrimination? And if so, is the interference due to the /s/ at all? Perhaps *any* preceding context would interfere with VOT perception.

Experiment 3 examined these two questions. Like Experiment 2, it included three discrimination conditions, one in which the test stimuli occurred in isolation and two in which they were preceded by a carrier phrase. In one instance, the carrier phrase was unambiguously *Take this*, while in the other it was *Take the*.

Methods

Subjects

Thirteen new Yale undergraduates were paid to participate. All were native speakers of American English.

Stimuli

The stimuli and test sequence in the baseline CVC condition were the same as in Experiment 2. In the *Take this*, or [-s]#CVC, condition, the original syllable-final [s] noise (94 ms long) and the original silent closure duration following it (115 ms long) were reinstated, which made this context quite unambiguous. A new carrier phrase was recorded for the *Take the* condition. To avoid closure voicing, it was produced as *Take the pin* by the same female speaker, with a silent closure duration of 84 ms. The stimuli from the *bin-pin* continuum were substituted for the original *pin*.

Procedure

As in Experiments 1 and 2, the baseline condition was always presented first, and the order of the two precursor discrimination conditions was counterbalanced across subjects.

Results

The results are shown in Figure 3. The CVC condition again yielded a pronounced peak in the short VOT range. This time, both precursor conditions also showed peaks, but they were somewhat lower and shifted towards longer VOTs. The peak for the *Take the* precursor stimuli was at a longer VOT than that for the *Take this* precursor stimuli.[6]

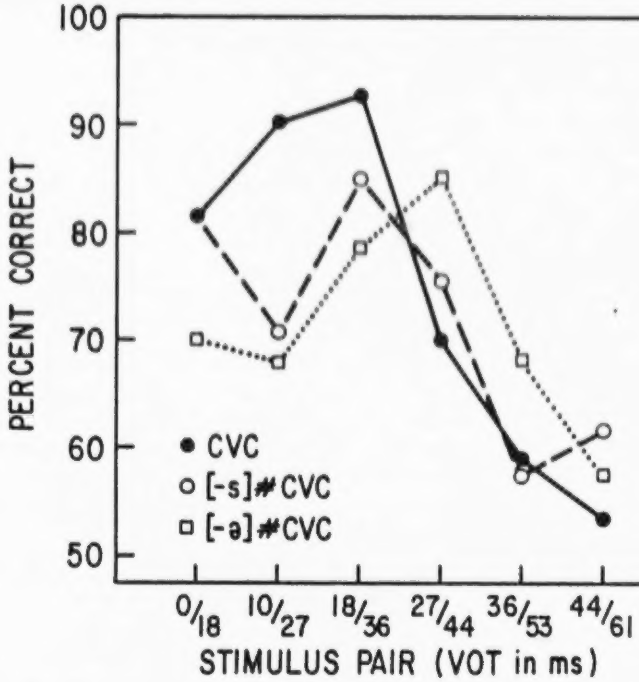


Figure 3. Discrimination performance as a function of stimulus pair in the three conditions of Experiment 3. The precursors are *Take this* (circles) and *Take the* (squares).

The ANOVA yielded a significant effect of stimulus pair, $F(5,60) = 31.01$, $p < .0001$, and a significant stimulus pair by condition interaction, $F(10,120) = 8.31$, $p < .0001$, but no significant main effect of condition. Comparing only the two precursor conditions, the two significant effects remained significant, $F(5,60) = 14.60$, $p < .0001$, and $F(5,60) = 4.56$, $p = .0014$, respectively. Thus there was reliable evidence only for shifts in peak location (which we will not attempt to explain here), not for a general performance decrement caused by precursors.

The results of this experiment, when compared with those of Experiment 2, show that an /s/ that unambiguously belongs to a preceding word interferes much less with VOT discrimination (if at all) than does an /s/ that has phonetic characteristics appropriate for word-initial position, everything else equal. This is entirely consistent with the phonological hypothesis. However, the auditory interference hypothesis is by no means ruled out. For one thing, the silent closure duration following the syllable-final [s] noise (itself a cue to a syllable boundary) was longer than that following the syllable-initial [s] noise. Naturally, this may have reduced any auditory interference. Then there were also acoustic differences between the two [s] noises in duration, amplitude envelope, and spectral detail that could have played a role. Moreover, the comparison between syllable-initial and syllable-final [s] noises was made across experiments, which is always problematic. Therefore, another experiment was conducted.

EXPERIMENT 4

The purpose of Experiment 4 was to assess independently the roles of [s] noise characteristics and silent closure duration in the precursor effect on VOT discrimination.

Methods

Subjects

Twelve Yale undergraduates, some of whom had participated also in Experiment 3, were recruited and paid for their services.

Stimuli

There were five conditions, three of which replicated those of earlier experiments: isolated CVC stimuli; CVC stimuli preceded by Take this, with the original 94-ms syllable-final [s] noise plus a 115-ms silent closure (as in Exp. 3); and CVC stimuli preceded by Take this, with the spliced-in 148-ms syllable-initial [s] noise plus a 77-ms closure (as in Exp 2). The two additional conditions represented the other two possible combinations of [s] noise and closure duration. Because of the increased number of conditions, the two extreme stimulus pairs (0/18 and 44/61 ms of VOT) were dropped to reduce test length, leaving only four two-step VOT contrasts. Each condition thus contained 16 different AXB triads, which were repeated five times.

Procedure

As usual, the CVC condition was presented first, and the order of the other four conditions was counterbalanced across subjects. The subjects were told that the precursor was always *Take this;* the phonetic differences among the precursor conditions were not mentioned in advance.

Results and Discussion

The results are shown in Figure 4. The discrimination function for the isolated CVC stimuli replicates that obtained in Experiments 2 and 3. The function for the precursor condition with syllable-final [s] and long closure also resembles that found in Experiment 3, showing but slight interference. The function for the precursor condition with syllable-initial [s] and short closure is quite different in shape from that found in Experiment 2, for unknown reasons. However, it does replicate the much greater interference obtained in that condition. The remaining two conditions, with mismatched [s] noises and closures, yielded results rather similar to the syllable-final [s] plus long closure precursor condition.

Alternatively, the data can be summarized as follows: All precursors interfered with VOT discrimination, though only at the shorter VOTs. Within the four precursor conditions, the combination of syllable-initial [s] and short closure led to much more interference than any of the other combinations. Thus, for the syllable-initial [s] noise, lengthening of the closure reduced interference considerably; for the syllable-final [s] noise, there was only a slight reduction. Similarly, at the short closure duration, a change in [s] noise made a large difference; at the long closure duration, only a small one.

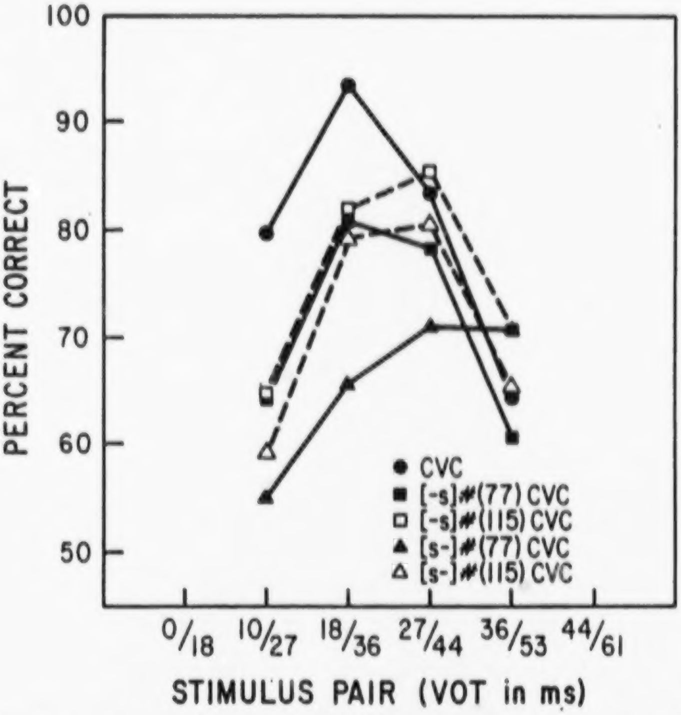


Figure 4. Discrimination performance as a function of stimulus pair in the five conditions of Experiment 4. The symbols [s-] and [-s] represent fricative noises from originally syllable-initial and syllable-final position, respectively. The numbers in parentheses are the closure durations. All precursors were intended to be perceived as *Take this.*

The statistical reliability of these effects was examined in two ANOVAs. The first analysis included only the four precursor conditions, with noise type, closure duration, and stimulus pair as factors. Apart from the expected main effect of stimulus pair, $F(3,33) = 12.79$, $p < .0001$, there were a significant main effect of noise type, $F(1,11) = 9.47$, $p = .0105$, a marginally significant main effect of closure duration, $F(1,11) = 4.85$, $p = .05$, a marginally significant interaction between closure duration and stimulus pair, $F(3,33) = 2.96$, $p = .0465$, and a significant triple interaction, $F(3,33) = 3.68$, $p = .0217$. The noise type by closure duration interaction was not significant. The triple interaction reflects the finding that the discrimination function in the long-noise/short-closure condition had a different shape than the functions in the other three conditions. The marginal significance levels indicate considerable variability among subjects.

A second ANOVA compared the isolated CVC condition with the precursor condition closest in performance level (syllable-final noise, long closure). The main effect of condition (i.e., the difference in average performance level) was not significant, but the stimulus pair by condition interaction (i.e., the difference in shape of the discrimination functions) was highly significant, $F(3,33) = 6.39$, $p = .0016$.

In summary, these results confirm the earlier finding that preceding [s] noises interfere with VOT discrimination as long as the VOTs compared are relatively short (40 ms or less), but not if they are relatively long. In addition, the results show that the interference depends both on [s] noise type and closure duration: The phonetic constellation appropriate for a syllable-initial /s/ leads to more interference at short

VOTs than any other noise-closure combination, thus substantially reducing the peak in the discrimination function.

These results are still compatible with both a phonological and an auditory interference account. From the perspective of the phonological hypothesis, they show that only the complete phonetic pattern characterizing syllable-initial /s/ leads to (overt or covert) /s/-stop cluster formation and phonological neutralization of the stop voicing contrast. From the auditory perspective, the two types of [s] noise generated different amounts of auditory interference because of their different acoustic properties (duration, amplitude, etc.), and this differential interference was reduced at longer temporal separations because of a ceiling effect on discrimination performance.

In a parallel study, we (Repp & Lin, 1989, Exp. 1) collected identification data for the very same stimuli, with five different closure durations ranging from 45 to 150 ms. Subjects labeled the stop consonants as "B" or "P" less accurately when the syllable-initial [s] noise preceded them, with a strong bias towards "P" responses, than when the syllable-final [s] preceded them. This difference decreased only slightly as closure duration increased and was still present at the longest closure. This pattern diverges from the present discrimination results, which already show a close convergence at a closure duration of 115 ms. Thus, as closure duration increased, discrimination performance exceeded what would be predicted on the basis of phonemic labeling in the syllable-initial [s] precursor condition. Here is a suggestion that the category boundary effects in that condition, at least, did not derive directly from attention to phonological categories, though it is also possible that covert labeling in the AXB task did not follow the same pattern as overt labeling in the identification test (cf Repp, Healy, & Crowder, 1979).

As a final attempt to distinguish between the two alternative accounts of the precursor interference effects, Experiment 5 returned to the method of nonspeech analogs, in defiance of its inherent problems.

EXPERIMENT 5

In this experiment, the entire precursors of Experiment 4 were replaced with broadband noises having exactly the same durations, overall amplitudes, and amplitude envelopes (cf. Salasoo & Pisoni, 1985; Gordon, 1988). Only spectral structure was eliminated. Thus, Experiment 5 also partially replicated Experiment 1, where a more primitive kind of nonspeech noise precursor had been used. To shorten the experiment, only the short-closure conditions of Experiment 4 were included. The prediction was that, if the different amounts of interference generated by the two kinds of [s] noise in Experiment 4 were due to differences in their acoustic properties (other than spectral differences), then the two nonspeech noises likewise should generate different amounts of interference, similar to those produced by the [s] noises. If, on the other hand, the difference between the two [s] noise conditions in Experiment 4 was due to spectral or specifically phonetic factors (i.e., /s/-stop cluster formation at some prelexical level in perception), then the two nonspeech noises should have equivalent effects, similar in magnitude to the effect of the syllable-final [s] noise in Experiment 4 or even smaller.

Methods

Subjects

Twelve subjects from the same pool participated. Some of them had also been subjects in Experiment 4.

Stimuli

Each of the two entire *Take this* precursors was converted into envelope-matched broadband noise using a computerized procedure first described by Schroeder (1968), which randomly reverses the polarity of digital sampling points with a probability of 0. 5. The resulting noise has a flat spectrum but retains the amplitude envelope of the speech.[7] It thus sounds vaguely speechlike but is not identifiable as an utterance. The stimuli from the *bin-pin* continuum were presented in isolation and preceded by either of the two noise precursors, with intervening silent intervals of 77 ms. The stimulus sequences were the same as those of the corresponding conditions in Experiment 4.

Procedure

As in previous experiments, the two precursor conditions were presented after the isolated CVC condition, in counterbalanced order. The subjects were told that the words would be preceded by a noise, which they should ignore.

Results and Discussion

The results are shown in Figure 5. It can be seen that the noise precursor derived from *Take this* with syllable-final [s] interfered minimally with VOT discrimination, but that the other precursor, which had the amplitude envelope of *Take this* with syllable-initial [s], did reduce performance at the shorter VOTs. This pattern was quite similar to that obtained with the corresponding speech precursors, though the absolute amount of interference was less. There was also considerable variability among subjects. In the ANOVA including all three conditions, there was a significant main effect of condition, $F(2,22) = 3.87$, $p = .0364$, and a significant condition by stimulus pair interaction, $F(6,66) = 2.42$, $p = .0354$, both of which were obviously due to the more effective precursor condition; the stimulus pair main effect was, as always, highly significant.

These results suggest that the different amounts of interference caused by syllable-final and syllable-initial [s] noises (Exp 4) are at least partially due to acoustic differences between the two noises. Since spectral differences were eliminated in the nonspeech precursors, and the duration differences between the original [s] noises were less well defined in the nonspeech analogs because of the absence of spectrally marked segment boundaries, this leaves differences in absolute amplitude and amplitude contour at noise offset as possible factors. This weakens the phonological account of the differences observed in Experiment 4. Also, the possibilities (mentioned in connection with Experiment 1) of hearing the nonspeech noise as a fricative or perceptually restoring a hidden fricative noise seem less plausible here, where the whole precursor phrase was transformed into noise. As in Experiment 1, however, nonspeech noise (Exp 5) interfered less with VOT discrimination than did [s] noise (Exp 4). This may also reflect acoustic differences—viz., the different spectral composition of the noises. Why a noise with predominantly high-frequency components (the natural [s]) should interfere more with the auditory processing of VOT than a broadband noise is not clear, however. Alternatively, the difference may have been caused by a reduced perceptual coherence of the nonspeech precursors with the following speech, compared to all-speech stimuli. A similar explanation was proposed by Gordon (1988) when he failed to find an effect of amplitude-modulated noise precursors on the perception of speech stimuli differing in closure duration. This explanation presumes that part or all of the interference takes place at an auditory level beyond the periphery, where the allocation of perceived sources plays a role. An acoustic factor disrupting source continuity may have been the presence of relatively strong low-frequency aperiodic energy in the nonspeech precursors, which was absent from the following speech.

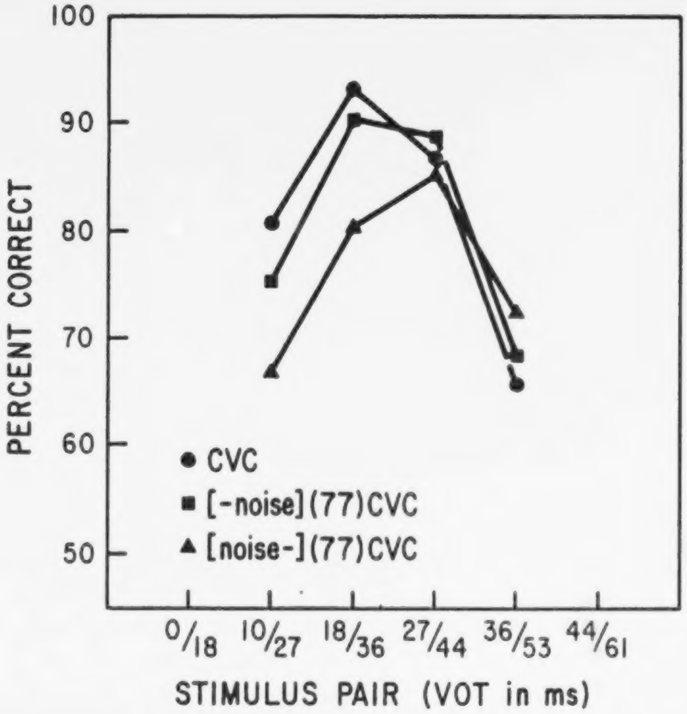


Figure 5. Discrimination performance as a function of stimulus pair in the three conditions of Experiment 5 The [-noise] precursor was derived from *Take this* with syllable-final [s], and the [noise-] precursor was derived from *Take this* with (spliced-in) syllable-initial [s] Filled symbols are used to match the corresponding speech precursor conditions in Figure 4.

GENERAL DISCUSSION

The present series of experiments started with an attempt to suppress the category boundary effect on a VOT continuum by preceding the stimuli with /s/ and thus neutralizing the phonological voicing contrast (Experiment 1). This manipulation was highly successful in that the discrimination peak indeed disappeared. However, a control condition with a nonspeech noise precursor also yielded some interference. Experiment 2 showed that the interference caused by an [s] noise was not affected by whether a word boundary preceded or followed the /s/. Experiment 3 suggested that an [s] noise with syllable-initial phonetic properties interferes more than one with syllable-final properties, and this was confirmed in Experiment 4, though only when the intervening closure duration was relatively short. Experiment 5 indicated that this difference was due to acoustic differences among the [s] noises, since amplitude-matched nonspeech noise precursors showed a similar difference, though less interference in absolute terms.

The results of several of these experiments could be interpreted as lending support to the hypothesis that the VOT discrimination peak (the "category boundary effect") originates at a phonological level of speech processing, not at the level of basic auditory sensitivities. Kewley-Port et al. (1988) recently arrived at the same conclusion when they observed that the discrimination peak was absent in a low-uncertainty discrimination task with trained subjects. Their conclusion has been challenged, however, by Macmillan et al. (1988) who did find a discrimination peak in a similar experiment that sampled the VOT continuum more finely and concluded that there is a psychoacoustic boundary on a VOT continuum. The present results are consistent with that interpretation as well. This ambiguity of interpretation pervades Experiments 1-4, but Experiment 5 tends to favor a psychoacoustic account of the precursor effects studied here. That is, the results suggest that the effect of a preceding /s/ on discrimination performance was caused not so much (or not at all) by the neutralization of the phonological voicing contrast in the following stop consonant as by interference with the auditory processing of VOT. This interference may then be considered a possible reason for why phonological neutralization of voicing contrasts in /s/-stop clusters is common in the languages of the world. (See Footnote 1, however.)

The mechanism of this interference is not known. It would require a whole series of further studies to disentangle the many possibilities. Voice onset time discrimination may rely not only on purely temporal differences but also on differences in F_1 onset frequency (Soli, 1983), in the relative strength of aspiration (Repp, 1979), and in the amplitude envelope at voicing onset (Darwin & Seton, 1983). A preceding noise could interfere with the processing of any or all of these. How such interference could result in the complete elimination of the psychoacoustic boundary around 20 ms of VOT is still not clear.

One way in which a noise precursor might affect auditory processing is through peripheral forward masking or adaptation. It is not clear, however, why adaptation of nerve fibers sensitive to the high frequencies characteristic of [s] noises should interfere with the perception of spectral and temporal signal properties in the low-frequency region, which VOT discrimination mainly relies on (voicing onset, F_1 onset frequency). A more plausible interpretation is that the presence of a precursor simply increased the complexity of the stimuli and thus made it more difficult for listeners to focus on the acoustic properties relevant to the task. This interference then may have taken place largely in auditory memory, rather than in peripheral auditory processing. This hypothesis is supported by the finding that nonspeech noise precursors generally interfered less with VOT discrimination than did speech precursors. Although speech and nonspeech precursors differed in spectral content and thus were not fully equated in their acoustic properties, the main difference between them may have been that the speech precursor "cohered" with the following word while the nonspeech precursor did not to the same extent. A parsing of the auditory input into likely sources probably precedes storage in auditory memory (cf. Bregman, 1978; Crowder, 1983), and a listener's knowledge of what constitutes a likely speech source may influence this parsing.

ACKNOWLEDGMENT

This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. We are grateful to Leigh Lisker and Neil Macmillan for helpful comments on an earlier version of the manuscript.

REFERENCES

Bregman, A. S. (1978). The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII* (pp. 63-76). Hillsdale, NJ: Erlbaum.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62,* 961-970.

Christie, W. M., Jr. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America, 55,* 819-821.

Cole, R. A., & Scott, B. (1973). Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology, 27*, 441-449.

Crowder, R. G. (1983). The purity of auditory memory. *Philosophical Transactions of the Royal Society* (London), *B302*, 251-265.

Darwin, C. J., & Seton, J. (1983). Perceptual cues to the onset of voiced excitation in aspirated initial stops. *Journal of the Acoustical Society of America, 74*, 1126-1135.

Diehl, R. L., Kluender, K. R., & Parker, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception and Performance, 11*, 209-220.

Dooling, R. J., Okanoya, K., & Brown, S. D. (1989). Speech perception by budgerigars (*Melopsittacus undulatus*): The voiced-voiceless distinction. *Perception & Psychophysics, 46*, 65-71.

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics, 43*, 137-146.

Howell, P., & Rosen, S. (1985). Natural auditory sensitivities as universal determiners of phonemic contrasts. In B. Butterworth, B. Comrie, & O. Dahl (Eds..), *Explanations of linguistic universals* (pp. 205-235). The Hague: Mouton.

Kewley-Port, D., Watson, C. S., & Foyle, D. C. (1988). Auditory temporal acuity in relation to category boundaries; speech and nonspeech stimuli. *Journal of the Acoustical Society of America, 83*, 1133-1145.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America, 70*, 340-349.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phoneme boundaries for the voicing feature in macaques. *Perception & Psychophysics, 32*, 542-550.

Lisker, L. (1984). How is the aspiration of English /p,t,k/ 'predictable'? *Language and Speech, 27*, 391-394.

Macmillan, N. A., Goldberg, R. F., & Braida, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America, 84*, 1262-1280.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops. *Journal of the Acoustical Society of America, 61*, 1352-1361.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech, 27*, 173-189.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice*, (Vol. 10, pp. 243-335). New York: Academic Press.

Repp, B. H. (1985). Perceptual coherence of speech: Stability of silence-cued stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 799-813.

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 129-145.

Repp, B. H., & Liberman, A. M. (1987). Phonetic category boundaries are flexible. In S. N. Harnad (Ed.), *Categorical perception* (pp. 89-112). New York: Cambridge University Press.

Repp, B. H., & Lin, H.-B. (1989). *Effects of preceding context on the voice-onset-time boundary*. Unpublished manuscript.

Rosen, S., & Howell, P. (1987a). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. N. Harnad (Ed.), *Categorical perception* (pp. 113-160). New York: Cambridge University Press.

Rosen, S., & Howell, P. (1987b). Is there a natural sensitivity at 20 ms in relative tone-onset-time continua? A reanalysis of Hirsh's (1959) data. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 199-209). Dordrecht: Martinus Nijhoff.

Sachs, R. M., & Grant, K. W. (1976). Stimulus correlates in the perception of voice onset time (VOT): II Discrimination of speech with high and low stimulus uncertainty. *Journal of the Acoustical Society of America, 60* (Suppl. No. 1), S91. (Abstract)

Salasoo, A., & Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language, 24,* 210-231.

Samuel, A. G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics,* 22, 321-330.

Sawusch, J. R., & Jusczyk, P. (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 408-421.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 44,* 1735-1736.

Soli, S. D. (1983). The role of spectral cues in discrimination of voice onset time differences. *Journal of the Acoustical Society of America, 73,* 2150-2165.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review, 77,* 234-249.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin, 96,* 371-383.

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics, 21,* 289-297.

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America, 60,* 1381-1389.

FOOTNOTES

[*] *Perception & Psychophysics, 45,* 323-332 (1989).

[†] Also, University of Connecticut, Storrs. Now at the Lexington Center, Jackson Heights, NY.

[1] It should be kept in mind that these predictions concern the discriminability of small VOT differences in the region of the original discrimination peak. Few listeners would fail to discriminate the extreme tokens of [spa] and [spʰa], despite the phonological neutralization.

[2] This noise was excised from a male speaker's production of the word *spectacular* in a sentence context. This rather short noise was used for no better reason than that it happened to be readily available in digitized form when the stimuli were constructed. As will be seen, however, it served its purpose well.

[3] This variability was apparently not due to effects of test order, which were nonsignificant in a separate analysis. That analysis also employed an arcsine transformation of the response proportions, which left the pattern of the results unchanged. Subsequent analyses did not include these two refinements.

[4] In the abbreviations for the conditions, the number sign (#) stands for a linguistic word boundary, and the dash (-) following the [s] represents the fact that the [s] noise had syllable-initial phonetic properties.

[5] The shift in the peak to a longer VOT value relative to Experiment 1 could be due to any of the many acoustic differences between the synthetic and natural stimuli: presence versus absence of a release burst, different amplitude envelopes, different vowels, different syllable structure and duration. It is well known that the VOT category boundary does not "stand still" but is influenced by a variety of extraneous variables (see Repp & Liberman, 1987).

[6] A corresponding difference in phoneme boundaries for the same stimuli was obtained in a labeling test administered to the same subjects. This boundary shift became the subject of a separate investigation (Repp & Lin, 1989) and will not be discussed further here. Suffice it to note that it cannot have been due to coarticulatory cues to the original following /p/ in the *Take the* precursor, because this should have caused a boundary shift in the opposite direction. Repp and Lin (1989) also showed that the difference in closure duration was not responsible.

[7] Actually, the noises presented to the subjects had a sloping rather than a flat spectrum. This was because the speech had been digitized with high-frequency pre-emphasis (of about 6 dB per octave above 1 kHz), and the digital noise files and the speech had to be (re)converted into sound in the same stimulus sequence, for which the computer demanded compatible specifications. Thus, even though

the nonspeech noises in the computer files had a flat spectrum, high-frequency de-emphasis was applied at output. This could have been circumvented by first re-digitizing the speech without pre-emphasis, but it was not considered enough of a problem to warrant the effort.

Can Speech Perception be Influenced by Simultaneous Presentation of Print?*

Ram Frost,[†] Bruno H. Repp, and Leonard Katz[††]

When a spoken word is masked by noise having the same amplitude envelope, subjects report they hear the word much more clearly if they see its printed version at the same time. Using signal detection methodology, we investigated whether this subjective impression reflects a change in perceptual sensitivity or in bias. In Experiment 1, speech-plus-noise and noise-only trials were accompanied by matching print, nonmatching (but structurally similar) print, or a neutral visual stimulus. The results revealed a strong bias effect: The matching visual input apparently made the amplitude-modulated masking noise sound more speechlike, but it did not improve the detectability of the speech. However, reaction times for correct detections were reliably shorter in the matching condition, suggesting perhaps subliminal facilitation. The bias and reaction time effects were much smaller when nonwords were substituted for the words, and they were absent when white noise was employed as the masking sound. Thus it seems that subjects automatically detect correspondences between speech amplitude envelopes and printed stimuli, and they do this more efficiently when the printed stimuli are real words. This supports the hypothesis, much discussed in the reading literature, that printed words are immediately translated into an internal representation having speechlike characteristics.

In the process of recognizing spoken words the listener must generate from the acoustic signal an internal representation that can make contact with the entries in the mental lexicon. A question of great importance for contemporary theories of speech perception is whether or not the generation of that representation is independent of lexical processes. One possibility is that the perceptual analysis of the speech input is completed before any contact with the mental lexicon occurs. Alternatively, some or all stages of the perceptual analysis may be interactively influenced by lexical processes that have been set in motion by partial information, prior context, or expectations. (See Frauenfelder & Tyler, 1987, for a review.)

Researchers concerned with auditory word perception generally take it for granted that the representations of words in the mental lexicon are phonologic in nature. (A notable exception is Klatt, 1980.) Investigators of visual word perception, too, often assume that phonological representations are accessed, although sometimes they postulate the existence of a separate visual-orthographic lexicon. While there are results indicating rapid visual recognition of written words prior to phonological analysis in certain tasks, there is much evidence that reading involves a phonological lexicon at some stage. (See McCusker, Hillinger, & Bias, 1981, for a review.) Theoretical parsimony dictates that this lexicon be the same as the one accessed in auditory word recognition. If so, then the process of speech perception might be penetrable visual influences (as well as the reverse): If earlier stages of auditory word perception can be affected by lexical processes, and if those same lexical processes can be activated in parallel by a visual presentation of print, then perception of words in the auditory modality could be influenced by words presented in the visual modality.

Evidence for lexical top-down effects within the auditory modality has been obtained in tasks involving phoneme restoration (Samuel, 1981; Warren, 1970), detection of mispronunciations (Cole & Jakimik, 1980), and shadowing (Marslen-Wilson & Welsh, 1978). It is not known, however, whether similar effects on speech perception can be elicited by visual input. That visual and auditory speech information can interact at a rather early level in perception has been demonstrated by McGurk and MacDonald (1976): The visual presentation of articulatory gestures (a speaker' s face) can affect subjects' perception of speech segments, even when the auditory input is unambiguous. However, because speech gestures are fundamental correlates of phonetic categories, their effect on speech perception takes place even before phonetic categorization, and certainly before lexical access (see Summerfield, 1987). The mapping of print into speech is far less direct and must be mediated by a lexical phonological level. Can a simultaneous presentation of printed words nevertheless influence the perception of speech?

A recent study by Frost and Katz (1989) suggests that it might. These authors presented printed and spoken words simultaneously and asked subjects to judge whether the words were the same or different. The experiment included a condition in which the speech was degraded severely by added signal-correlated noise (a broadband noise with the same amplitude envelope as the stimulus). Nevertheless, the subjects found the task fairly easy; the average error rate was only 10%. In a subsequent pilot study, the same authors presented the subjects simultaneously with both degraded speech *and* degraded print. Here, subjects' performance was close to chance. The subjects' phenomenological description was that they often could not hear any speech at all in the auditory input, whereas previously, when clear print matching a degraded auditory word was presented simultaneously, they reported no difficulty in identifying the degraded word. Thus, it seemed as if the presence of the printed word enabled the subjects to separate the speech from the noise, and hence to perceive it much more clearly. There is another possibility, however: Subjects' introspections may have reflected merely an illusion caused by the correspondence between the print and the amplitude envelope of the masking noise, which was identical with that of the speech. That is, subjects might have thought they heard speech even if the masking noise alone had been presented accompanied by "matching" print; however, no such trials were included.

The following experiments were conducted to determine whether simultaneous presentation of a printed word can truly facilitate the detection of a speech signal in noise. A positive answer to this question would provide strong support for the hypothesis that visual and auditory word perception are functionally interdependent. On the other hand, even the finding of a pure "response bias" elicited by the correspondence between the print and the noise amplitude envelope would be interesting, as it, too, represents an effect of print on speech perception, though of a different kind.

EXPERIMENT 1

In Experiment 1, subjects were presented with either speech plus noise or with noise alone, and their task was to detect the presence of the speech signal. In conjunction with the auditory presentation, the subjects saw on a computer screen a matching word, a nonmatching word, or a neutral stimulus (XXXXX). The purpose of the experiment was, first, to confirm our pilot observations that simultaneous presentation of matching print makes subjects hear speech in the noise and, second, to examine whether that effect reflects an increase in sensitivity to speech actually present, or a bias to interpret the masking noise as speech. The signal detection paradigm we used enabled us to compute independent indices of sensitivity and bias.

It is important to keep in mind that even on noise-only trials there could be a (partial) match between the visual and auditory stimuli, since each spoken word had its individual envelope-matched masking noise. Another way of framing the question, therefore, was to ask whether matching print would enhance the detectability of the *spectral* features of speech hidden in the noise, or whether it would have an effect on the subjects' responses because the auditory *amplitude envelope* corresponds to that of the spoken form of the printed word. These two effects are not mutually exclusive and may operate simultaneously.

Methods

Subjects

Thirty-six undergraduate students, all native speakers of English, participated in the experiment for payment.

Stimulus preparation

The stimuli were generated from 24 regular, disyllabic English words that had a stop consonant as their initial phoneme. All words were stressed on the first syllable. The number of phonemes in each word ranged from four to six, and the word frequencies, according to Kučera and Francis (1967), ranged from 0 to 438, with a median of 60. The words were spoken by a female speaker in an acoustically shielded booth and were recorded on an Otari MX5050 tape recorder. They were then digitized at a 20 kHz sampling rate. From each digitized word, we created a noise stimulus with the same amplitude envelope by randomly reversing the polarity of individual samples with a probability of 0.5 (Schroeder, 1968). Such signal-correlated noise retains a certain speech-like quality, even though its spectrum is flat and it cannot be identified as a particular utterance unless the choices are very limited (see Van Tasell, Soli, Kirby, & Widin, 1987). The speech-plus-noise stimuli were created by adding the waveform of each digitized word to that of its matched noise, applying scaling factors to vary the speech-to-noise (S/N) ratio while keeping the overall amplitude constant. Six different S/N ratios were used: -9.5 dB, -10.7 dB, -12 dB, -13.2 dB, -14.4 dB, -16.5 dB. All these ratios were well below the identification threshold, according to earlier observations. (A ratio of -7.5 dB was used in the pilot study referred to in the introduction.) Because each word had its own amplitude-matched masking stimulus, any given S/N ratio was exactly the same for all words and even for different phonetic segments within each word.

The visual stimuli were presented on a Macintosh computer screen in bold face Geneva font. They subtended an average visual angle of approximately 2.5 degrees. Their presentation was triggered by a tone recorded at the onset of the auditory stimulus, on a second audio channel. The onsets of the spoken words were determined visually on an oscilloscope and were verified auditorily through headphones. The onset was defined as the release of the initial stop consonant in all cases.

Design

There were three experimental groups of twelve subjects each. Each subject was tested at two of the six different S/N ratios: relatively high (-9.5 and -10.7 dB), medium (-12 and -13.2 dB), or low (-14.4 and -16.5 dB). At each S/N ratio there were 144 trials. Each of the 24 noise and 24 speech-plus-noise stimuli was presented in three different visual conditions: (1) a matching condition (i.e., the same word that was presented auditorily, and/or that was used to generate the noise, was presented in print), (2) a nonmatching condition (i.e., a different word having the same number of phonemes and a similar phonological structure as the word that was presented auditorily, or was used to generate the noise, was presented in print; e.g. PERSON—BASKET), and (3) a neutral condition in which the visual stimulus was XXXXX.

Procedure and apparatus

The subject was seated in front of the Macintosh computer screen and listened binaurally over Sennheiser headphones. The task consisted of pressing a "yes" key if speech was detected in the noise, and a "no" key if it was not. The dominant hand was used for the "yes" responses. Although the task was introduced as purely auditory, the subjects were requested to attend carefully to the screen as well. They were told in the instructions that, when a word was presented on the screen, it was sometimes similar to the speech or noise presented auditorily, and sometimes not. However, they were informed about the equal proportions of "yes" and "no" trials in each of the different visual conditions.

The tape containing the auditory stimuli was played on a two-channel Crown 800 tape recorder. The verbal stimuli were transmitted to the subject' s headphones through one channel, and the trigger tones were transmitted through the other channel to an interface that directly connected to the Macintosh, where they triggered the visual presentation and the computer' s clock for reaction time measurements.

The experimental session began with 24 practice trials, after which the first 144 trials were presented in one randomized block, starting with the higher S/N ratio. Then there was a three-minute break before the second, more difficult block employing the lower S/N ratio.

Results and Discussion

Response percentages

For each subject we determined the percentages of "yes" and "no" responses to speech-plus-noise and noise-only stimuli in each of the three visual conditions. Table 1 shows the average percentages of "yes" responses (i.e., of hits and false alarms); the percentages of "no" responses (misses and correct rejections, respectively) are their complements. There was an extremely high rate of false alarms in all conditions, due to the speechlike envelope of the signal-correlated noise. It is evident that hits decreased and false alarms increased with decreasing S/N ratio, as expected. Most interestingly, we see that the percentage of correct detections was higher in the matching print condition than in the other two conditions. This replicates the pilot observations that led to the present experiment. However, the percentage of false alarms was highest in the matching condition also. Apart from the issue of statistical reliability, this raises the question of whether we are dealing here with an increased bias to say "yes" in the matching condition, regardless of whether spectral features were present or not, or whether detectability of spectral properties of speech was in fact increased in the matching condition. To address this question, we turn to an examination of independent discriminability (or sensitivity) and bias indices.

TABLE 1. Percentages of hits and false alarms (Exp. 1).

S/N Ratio	Hits			False alarms		
	Match	Nomatch	XXX	Match	Nomatch	XXX
-9.5 dB	96	92	93	37	19	14
-10.7 dB	97	92	90	41	26	17
-12.0 dB	90	77	77	43	27	21
-13.2 dB	89	80	77	48	29	32
-14.4 dB	74	61	51	56	44	33
-16.5 dB	72	58	45	59	46	34
Average	86	77	72	47	32	25

Discriminability and bias indices

Indices of discriminability and bias were computed following the procedures of Luce (1963). Luce's indices were preferred over the standard measures of signal detection theory, d' and Beta, because they are easier to compute and do not require any assumptions about the shapes of the underlying signal and noise distributions. Moreover, earlier comparisons have shown that results couched in terms of signal detection and Luce indices tend to be very similar (see, e.g., Wood, 1976). The Luce indices, originally named -ln(Eta) and ln(b), but renamed here for convenience *d* and *b*, respectively, are:

$$d = (1/2)\ln[p(\text{yes} \mid S{+}N)p(\text{no} \mid N)/p(\text{yes} \mid N)p(\text{no} \mid S{+}N)]$$

and

$$b = (1/2)\ln[p(\text{yes} \mid S{+}N)p(\text{yes} \mid N)/p(\text{no} \mid S{+}N)p(\text{no} \mid N)]$$

where S+N and N stand for speech-plus-noise and noise alone, respectively. The discriminability index *d* assumes values in the same general range as the d' of signal detection theory, with zero representing chance performance. The bias index *b* assumes positive values for a tendency to say "yes" and negative values for a tendency to say "no."[1]

The average indices are shown in Table 2. Each index was subjected to a three-way analysis of variance with the factors subject group (actually, S/N ratio between groups), S/N ratio (within groups), and visual condition. The *d* indices confirm that subjects' performance deteriorated as the S/N ratio decreased. At a S/N ratio of -16.5 dB, performance was almost at chance level. The main effect of subject group was significant, $F(2,33) = 24.8$, $p < 0.001$, though the main effect of S/N ratio within subject groups was not. The latter finding probably represents a practice effect: The more difficult S/N ratio always came last in the experimental session and thus received the benefits of practice. The most important result, however, is that subjects' sensitivity was *not* increased in the matching condition. On the contrary, the average *d* index was lowest in that condition and highest in the neutral condition, though the main effect of visual condition was not significant, $F(2,66) = 1.53$, $p = 0.2$. These differences among visual conditions seemed to be reliable at the two highest S/N ratios only, as suggested by a significant interaction of visual condition and subject group, $F(2,66) = 3.21$, $p = 0.02$. We have no explanation for this finding at present. However, our hypothesis that simultaneous matching print might facilitate the detection of the spectral features of speech in noise is clearly disconfirmed.

TABLE 2. Discriminability (*d*) and bias (*b*) indices (Exp. 1).

S/N Ratio	*d*			*b*		
	Match	No match	XXX	Match	Nomatch	XXX
-9.5 dB	2.06	2.36	2.68	1.36	0.27	0.36
-10.7 dB	2.05	2.21	2.32	1.51	0.63	0.22
-12.0 dB	1.42	1.38	1.53	1.06	0.10	-0.04
-13.2 dB	1.46	1.55	1.38	1.24	0.21	0.19
-14.4 dB	0.48	0.44	0.40	0.70	0.05	-0.45
-16.5 dB	0.37	0.30	0.23	0.87	0.09	-0.52
Average	1.30	1.37	1.42	1.13	0.22	-0.04

Turning now to the bias indices, we see a striking difference among the visual conditions: Overall, there was a strong tendency to say "yes" in the matching condition, but little or no bias in the other two conditions. The main effect of visual condition was highly significant, $F(2,66) = 57.1$, $p < 0.001$. In addition, it appears that the overall frequency of "yes" responses decreased with S/N ratio in all visual conditions, but this tendency did not reach significance, due to considerable between-subject variability.

The increased frequency of "yes" responses when matching print was present, without a concomitant increase in signal detectability, was obviously caused by the speechlike qualities of the masking noise. In the matching condition, the amplitude envelope of the noise was appropriate for a spoken version of the printed word, and therefore it seemed to the subjects that the word was presented auditorily, whether or not it was in fact hidden in the noise. In retrospect, this explains the subjective impressions of "hearing" words in noise accompanied by print, that led to the present series of experiments. Our data thus reveal that subjects, even when they are not explicitly instructed to do so, automatically detect the correspondence between the amplitude envelope of a nonspeech signal and a sequence of printed letters forming a word, with the consequent illusion of actually hearing the word. This illusion seems akin to the phoneme restoration phenomenon, where surrounding speech context leads subjects to "hear" single phonemes whose acoustic correlates have been replaced by some suitable masking noise (Samuel, 1981; Warren, 1970). The effect revealed in our research suggests that *all* the phonemes in a word may be restored, at least to some extent, when the speech amplitude envelope carried by noise is accompanied by matching print. Since there cannot be a direct connection between the printed letters and the auditory amplitude contour, this might be a top-down effect mediated by a speechlike internal representation of the printed word. Although a global phonetic representation could be envisioned that contains envelope information without a more detailed segmental coding, the striking difference between the matching and nonmatching visual conditions suggests otherwise: The nonmatching printed words were in fact fairly similar to the matching ones in syllabic stress pattern and phonologic structure, so that a detailed knowledge of the segmental structure would seem to have been necessary to discriminate between their envelopes (see Van Tasell et al., 1987). We conclude, therefore, that printed words are automatically transformed into a detailed phonetic representation, which is probably generated from a more abstract phonological representation stored in the mental lexicon.

Reaction times

Although measures of discriminability and reaction times are usually highly correlated, they may reflect different phases of the cognitive processes involved in the task. While discriminability indices tap into the conscious decision stage, latencies reflect subjects' confidence in reaching their decisions (see Luce, 1986). Therefore, an examination of reaction times may reveal additional information about subjects' processing of the stimuli.

In calculating the average latencies for each subject, outliers beyond two standard deviations from the mean were eliminated. Outliers accounted for less than two out of the 24 responses per condition, on the average. The average reaction times are presented in Table 3. At the higher S/N ratios, there were not enough misses ("no" responses on S+N trials) for meaningful averages to be calculated. Separate analyses of variance were conducted on hits, false alarms, and correct rejections.

Looking at the hits first, we see that the average latencies increased as the S/N ratio decreased across subject groups, $F(2,33) = 6.73$, $p = 0.003$. But no reliable decrease was found within subject groups, probably due to the aforementioned practice effect. As to the effect of visual presentation, we see that the average reaction times were some 100 ms faster in the matching condition than in the other two conditions. This difference was highly significant, $F(2,66) = 43.08$, $p < 0.001$, and extremely robust: Every single subject showed it, even at the lowest S/N ratios.

The false alarm latencies were significantly slower than the hit latencies across all S/N ratios, as confirmed in a separate comparison, $F(1,33) = 13.38$, $p = 0.001$. This is consistent with the common finding of slower reaction times for incorrect than for correct responses. However, there was no significant difference among the three visual conditions, nor was there a main effect of S/N ratio.

The correct rejection latencies, too, were slower than the hit latencies. They decreased with the S/N ratio within subject groups, $F(1,33) = 7.78$, $p = 0.009$, presumably due to practice. The magnitude of that decrease was largest at the lowest S/N ratios, which caused a subject group by S/N ratio interaction, $F(2,33) = 3.73$, $p = 0.03$. There was no difference between the matching and nonmatching conditions. However, reaction times were faster in the neutral condition. This effect of visual condition was quite consistent across different S/N ratios and was highly significant, $F(2,66) = 23.71$, $p < 0.001$.

The very reliable speeding up of hit responses in the matching visual condition could be explained in terms of the bias to respond "yes" in that condition. However, the false alarms did not show the same decrease even though they were subject to the same response bias (see Table 2). Also, correct "no" responses might have been expected to show longer latencies in the matching condition. Thus, the reaction time patterns of false alarms and of correct rejections suggest that it was not just the match of print and noise amplitude envelope that caused faster latencies for hits. Rather, it seems that spectral speech information had to be present in order for responses to be speeded up by a match. The faster hit latencies in the matching condition then may reflect, after all, an increase in subjects' sensitivity to the spectral features of the speech signal itself, even though overt detection was not enhanced, and even though the reaction time effect persisted at S/N ratios where detectability of the speech approached chance level. Thus, the latencies may tap an earlier level of processing that preceded the conscious decision about presence or absence of the speech signal. This would explain the absence of a similar effect for false alarms, because there was never any signal present for these responses. This interpretation remains speculative, however.

TABLE 3. Reaction Times (Exp. 1).

"YES" Responses						
S/N Ratio	Hits			False alarms		
	Match	Nomatch	XXX	Match	Nomatch	XXX
-9.5 dB	667	769	740	920	803	750
-10.7 dB	624	735	724	873	854	775
-12.0 dB	749	843	865	933	1057	1115
-13.2 dB	745	858	834	1015	876	948
-14.4 dB	910	1029	982	1003	1031	1094
-16.5 dB	838	967	951	874	1019	1054
Average	755	867	850	936	940	952

"NO" Responses						
	Misses			Correct rejections		
	Match	Nomatch	XXX	Match	Nomatch	XXX
-9.5 dB		(Insufficient data)		914	902	814
-10.7 dB				884	855	808
-12.0 dB				974	1024	896
-13.2 dB				1007	989	876
-14.4 dB	1125	1099	1024	1084	1124	987
-16.5 dB	961	988	904	913	977	909
Average				963	978	882

EXPERIMENT 2

The strong response bias caused by the match of print and noise amplitude envelope represents an influence of print on speech perception, a kind of "word restoration" illusion. The main purpose of Experiment 2 was to investigate whether this is a lexical or a pre-lexical influence. The phonetic representation generated from the print may have been derived from a phonological representation following lexical access or, alternatively, it may have been generated directly from the print *via* spelling-to-sound conversion rules. One possible method for distinguishing between these two alternatives is to present subjects with nonwords instead of words. Although some authors have argued that nonwords are pronounced by referring to related lexical entries for words (e.g., Glushko, 1979), this route is still less direct than that available for real words. Therefore, if the word restoration effect is lexical in origin, it should be reduced or absent for nonwords. If it is prelexical, on the other hand, it should be obtained for nonwords just as for words. In addition, we wondered whether the intriguing and extremely consistent reaction time facilitation for correct detection of words in the matching condition would be obtained for nonwords as well.

Methods

Subjects

Twelve undergraduate students, all native speakers of English, participated in the experiment for payment.

Stimulus preparation

The stimuli were generated from 24 disyllabic English pseudowords formed by altering one or two letters of real words having the same stress pattern. They had a stop consonant as their initial phoneme, and the number of phonemes ranged from four to six. The written and spoken forms of all nonwords exhibited a regular spelling-to-sound correspondence, according to Venezky (1970); that is, each printed nonword had only one plausible pronunciation—the one spoken. The method for constructing the auditory and the visual stimuli was identical to that of Experiment 1.

Design

Design, procedure, and apparatus of Experiment 2 were identical to those of Experiment 1, except that only one group of subjects was used. Each subject was tested at two S/N ratios: -12 dB and -14.4 dB, in this order.

Results and Discussion

Response percentages

The average percentages of hits and false alarms are presented in Table 4. When compared to the results obtained for words with the same S/N ratios (Table 1), it is clear that subjects' performance was worse with nonwords: The percentage of hits was lower, and the percentage of false alarms was higher. In addition, the effect of matching print was much smaller in the nonwords: The percentages of correct detections in the matching and nonmatching conditions were almost identical, and the false alarm percentages showed only a small difference. To examine these effects further we calculated the discriminability and bias indices.

TABLE 4. Percentages of hits and false alarms for nonwords (Exp. 2).

S/N Ratio	Hits			False alarms		
	Match	Nomatch	XXX	Match	Nomatch	XXX
-12.0 dB	75	69	64	47	41	33
-14.4 dB	68	69	61	53	47	39
Average	71	69	62	50	44	36

Discriminability and bias indices

The average *d* and *b* indices are presented in Table 5. The *d* indices show that subjects' performance deteriorated as the S/N ratio decreased. This main effect was significant, $F(1,11) = 9.3$, $p < 0.01$. At the higher S/N ratio, the *d* values were lower than those obtained for words in the previous experiment, suggesting that detection of nonwords was more difficult than that of words. At the lower S/N ratio, discriminability was low for both words and nonwords. Apparently, in the present experiment, subjects did not show any effect of practice. As with the words in Experiment 1, the different visual conditions did not affect subjects' sensitivity. The main effect of visual condition was nonsignificant, $F(2,22) = 0.09$.

TABLE 5. Discriminability (*d*) and bias (*b*) indices (Exp. 2).

S/N Ratio	*d*			*b*		
	Match	Nomatch	XXX	Match	Nomatch	XXX
-12.0 dB	0.86	0.73	0.76	0.63	0.24	-0.02
14.4 dB	0.39	0.51	0.56	0.51	0.34	-0.05
Average	0.63	0.62	0.66	0.57	0.29	-0.03

Analysis of the bias indices revealed a significant effect of visual condition $F(2,11) = 10.0$, $p < 0.001$. Although the direction of the effect was similar to that obtained for words, its size was much smaller for the nonwords. Moreover, a Tukey post-hoc analysis revealed that the bias indices in the matching and nonmatching condition did not differ significantly. In order to assess directly whether the bias effect in the three visual conditions interacted with the lexical status of the stimuli, we conducted a separate analysis in which the nonwords of Experiment 2, and the words of Experiment 1 (for comparable S/N ratios) were combined. The interaction of word/nonword and visual condition was significant, $F(2,92) = 6.97$, $p < 0.001$. This outcome demonstrates that the bias effect was indeed different for words and nonwords.

Reaction times

The average reaction times are presented in Table 6. The slow latencies, especially at the higher S/N ratio, suggest again that detection of nonwords was more difficult than detection of words. We conducted separate analyses for hits, false alarms, correct rejections, and misses. The pattern of the hits revealed no effect of visual presentation, $F(2,22) = 0.3$, in sharp contrast to the results for words. Thus, for nonwords, matching print did not facilitate correct "yes" responses. Also, neither the effect of S/N ratio nor the interaction of S/N ratio and visual condition were significant. The significance of the word-nonword difference was again assessed in a separate analysis in which data from Experiment 1 and 2 were combined. The interaction of word/nonword and visual condition was indeed significant, $F(2,92) = 3.27$, $p = 0.04$.

The false alarms analysis revealed no significant effect of visual presentation. The analysis of correct rejections, however, did show such an effect, $F(2,11) = 13.8$, $p < 0.001$, due to faster responses in the neutral condition. This unexplained effect is very similar to that found for words. The average reaction times for misses were relatively slow, without any significant effects.

In summary, in contrast to the results previously obtained for words, the bias to say "yes" in the matching condition (the "nonword restoration" effect) was much smaller, and reaction times for correct detections were not faster when the print matched the speech signal. These results support the hypothesis that the word restoration illusion is lexically mediated. Because nonwords are not represented in the mental lexicon, their covert pronunciation is either generated prelexically from the print, or indirectly by accessing similar words in the lexicon. Apparently, either process is too slow or too tentative to enable subjects to match the resulting internal phonetic representation to a simultaneous auditory stimulus before that stimulus is fully processed.

TABLE 6. Reaction Times (Exp. 2).

"YES" Responses						
S/N Ratio	Hits			False alarms		
	Match	Nomatch	XXX	Match	Nomatch	XXX
-12.0 dB	972	1019	1005	1149	1092	1245
-14.4 dB	1005	999	1017	1020	1071	1081
Average	988	1009	1011	1084	1082	1163
"NO" Responses						
	Misses			Correct rejections		
	Match	Nomatch	XXX	Match	Nomatch	XXX
-12.0 dB	1269	1265	1053	1129	1116	1002
-14.4 dB	1120	1070	1055	1048	1091	983
Average	1195	1167	1054	1088	1103	993

One unexpected finding was that overall performance was much worse for nonwords than for words, even though the stimuli were presented at exactly comparable S/N ratios. Had the task required identification of the stimuli, this difference would not have been surprising, since superior recognition performance for real words has been demonstrated in many studies of visual and auditory word perception. However, our subjects could not identify the masked speech, and in most cases they could hardly say if any speech was present at all. How, then, is the poorer performance with nonwords to be explained?

One possibility is that, because words are represented in the lexicon, they are better detected by the perceptual system. This hypothesis was disconfirmed in a recent study reported in detail elsewhere (Repp & Frost, 1988): When masked words and nonwords were randomly presented, without simultaneous print, they were detected equally well. Another possibility is that our subjects adopted different perceptual strategies in Experiments 1 and 2. The instantaneous presentation of the print occurred at the beginning of the speech, which unfolded over the next several hundred milliseconds. Almost certainly, processing of the print was completed before that of the speech. Because of this, and also because only words or nonwords were included in each of the experiments, the subjects always knew in advance whether the auditory stimulus was going to be a word or a nonword. We suspect that this foreknowledge was responsible for the observed differences in performance.

EXPERIMENT 3

The purpose of Experiment 3 was to examine the effects of matching print on detectability of words and nonwords in white noise, instead of signal-correlated noise. In white noise, the amplitude envelope fluctuates less, and randomly, rather than in a speechlike fashion. Therefore, the "word restoration" effect caused by the match of auditory amplitude envelope and print should disappear completely when white noise is used, and with it any difference in bias indices.

Even though signal detection theory treats sensitivity and bias as independent parameters, it is conceivable that, in the absence of a strong response bias due to speechlike noise, any effect of matching print on detectability of speech might emerge more clearly, particularly since both spectral and amplitude features can now be utilized by subjects for speech signal detection. Counteracting this possible advantage of using a white noise masker was the possibility that the separation of speech from white noise is much easier, and perhaps rests on more peripheral processing of the input. This in turn might reduce any potential top-down effects on subjects' sensitivity. Nevertheless, we wondered whether at least the reaction time difference found for words in Experiment 1 could be replicated in white noise.

Methods

Subjects

Twelve paid undergraduate subjects participated. All were native speakers of English.

Stimuli

The same words and nonwords as in Experiments 1 and 2 were used. To make the S/N ratio comparable for all stimuli, an individual white-noise masker was constructed for each speech stimulus as follows: First, white noise produced by a General Radio 1390-A random noise generator was sampled at a rate of 20 kHz and stored in a file. Next, a segment of exactly the same length as the speech was excerpted from that file. Then, 5-ms amplitude ramps were put at the beginning and end of the noise to avoid abrupt onsets and offsets. Subsequently, the average dB levels of the speech and of the white noise segment were determined, and the white noise (which, as recorded, was from 2 to 7 dB more intense than the speech) was attenuated digitally to exactly the same average amplitude as the speech. The S/N ratios, therefore, were specified relative to the average, not the peak, speech signal level. To obtain the speech-plus-noise stimuli, the speech and white noise waveforms were added digitally at a S/N ratio of -28 dB, keeping the overall amplitude constant. This ratio was based on previous data (Repp & Frost, 1988) and was intended to yield a level of performance around 75 percent correct.

Design and Procedure

Design, procedure, and apparatus were similar to those of Experiments 1 and 2, except that each subject was tested in both the word and the nonword conditions. Half the subjects received the word condition first, and half the nonword condition. The playback level was calibrated for each tape using white noise recorded at the beginning of the tape. The level was set so the calibration noise registered 0.1 V on a voltmeter, corresponding to 90 dB at the subjects' earphones. The average level of the stimuli was from 2 to 7 dB lower.

Results and Discussion

The average percentages of hits and false alarms are presented in Table 7. The results reveal that matching print did not increase either the hit rate or the false alarm rate for either words or nonwords.

TABLE 7. Percentages of hits and false alarms (Exp. 3).

Stimulus	Hits			False alarms		
	Match	Nomatch	XXX	Match	Nomatch	XXX
Words	78	75	69	21	21	21
Nonwords	78	78	74	22	22	14

Table 8 presents the *d* and *b* indices. The *d* indices for words and nonwords were very similar, without any significant effects of the different visual conditions. Overall performance in the experiment was relatively good, so the above findings do not result from the inability of subjects to detect the speech in noise.

Table 8 also reveals that, as predicted, the bias effect found in the previous experiments had disappeared. However, there was a significant tendency to give "no" responses in the neutral condition, for both words and nonwords $F(2,22) = 4.48$, $p = 0.02$. Apparently, the absence of a printed word influenced the subjects' decision criterion.

TABLE 8. Discriminability (*d*) and bias (*b*) indices (Exp. 3).

Stimulus	*d*			*b*		
	Match	Nomatch	XXX	Match	Nomatch	XXX
Words	1.36	1.30	1.25	0.02	-0.12	-0.38
Nonwords	1.34	1.32	1.63	-0.01	0.05	-0.45

The reaction times of hits and correct rejections are shown in Table 9. The numbers of false alarms and misses per subject were insufficient for statistical analysis. Hit latencies did not differ significantly across the different visual conditions, or between words and nonwords. This result is consistent with all of the above findings. RTs for correct rejections revealed a significant interaction of visual conditions and stimulus type $F(2,22) = 10.4$, $p < 0.001$. However, this interaction is uninterpretable because the "match" on noise-only trials concerned solely the average amplitude and duration of the noise. It seems unlikely that such "matches" had any influence on subjects' responses. The variability in reaction times may just reflect differential responses to the "matching" and "nonmatching" visual word stimuli.

TABLE 9. Reaction Times (Exp. 3).

	"YES" Responses			"NO" Responses		
Stimulus	Hits			Correct rejections		
	Match	Nomatch	XXX	Match	Nomatch	XXX
Words	706	731	717	818	891	804
Nonwords	720	722	762	909	882	857

In summary, as we hypothesized, when the masking noise did not include the envelope information, the effect of matching print on response bias disappeared. However, the detection of speech was not enhanced by matching print. Moreover, matching print did not affect reaction times of correct word detections, perhaps because the detection of speech in white noise rests on different criteria than detection in signal-correlated noise.[2] Speech detection in white noise may be a superficial task that does not require a detailed analysis of the auditory input, and therefore is "out of reach" for top-down effects. The absence of an overall performance difference between words and nonwords is also consistent with this interpretation.

GENERAL DISCUSSION

In the present study we used a signal detection task to investigate whether simultaneous presentation of matching print can affect the detectability of speech in noise. In Experiment 1 we found no evidence for an enhancement of subjects' sensitivity to the spectral features of the speech. However, the reaction time analysis revealed that subjects' confidence in their decisions was increased by the match, but only when spectral information was indeed present in the noise. There was also a strong bias toward "yes" responses, caused by the correspondence of print and noise amplitude envelope. From Experiment 2 we learned that printed nonwords elicit a much smaller bias effect and show no facilitation of reaction times. Finally, Experiment 3 demonstrated that there is no influence of print even for words when white noise is used as the masking sound.

The results of Experiment 1, together with earlier phenomenological observations, suggest that, when signal-correlated noise is employed, the presentation of matching print generates a perceptual illusion: Subjects believe they hear speech, even when no speech is present in the noise. That the bias is mediated by the amplitude envelope of the noise was confirmed in Experiment 3, where the effect was totally absent with white noise.

The bias effect suggests that the printed words were immediately recoded into an internal phonetic form. In order for subjects' responses to be affected by the match between print and noise amplitude envelope, the information generated from the visual and the auditory modalities must have been in the same internal metric. The amplitude envelope of the auditory stimulus was almost certainly insufficient to generate a detailed abstract phonologic code that could have been compared to phonologic information accessed from print. Therefore, it is the print that must have been converted internally into a phonetic representation. What our findings teach us is that a phonetic code is generated from printed words *automatically*, even when the task does not require it. After all, our subjects were never instructed to match the print to the auditory stimuli, and were specifically informed of the equal distribution of speech-plus-noise and noise-only trials in the different visual conditions. It is unlikely that amplitude envelopes are stored as such in the lexicon, since they are contingent on phonetic structure. Their availability from print implies that a segmental phonetic representation is generated. Thus, our results provide further confirmation for the notion of obligatory and fast phonetic coding in reading, suggested by a large literature on visual word perception but sometimes challenged by those who find evidence for rapid lexical access based on orthography alone. We suspect that phonetic coding takes place regardless of what information gets to the lexicon first.

How was it possible for matching print to influence reaction times only for correct detections, without an apparent increase in sensitivity to spectral speech features? A possible explanation of this pattern of results is that matching print had a confirmatory effect at a processing stage *following* the extraction of partial spectral features from the noise. Thus, the subjects' confidence in correct detections was increased in the matching condition without any actual increase in the amount of spectral information extracted.

The strong bias effect and the facilitation of reaction time were not obtained for nonwords. Therefore the influence of printed words on speech processing appears to be lexically mediated. That is, the internal phonetic representation is probably generated from a more abstract phonological code stored in the lexicon, rather than by applying spelling-to-sound conversion rules. Nonwords are clearly at a disadvantage in such process.[3] Whether this disadvantage consists merely of a longer processing time could be tested by presenting the visual information somewhat in advance of the onset of the auditory stimulus, so that there is sufficient time to recast nonwords into an internal phonetic code. The observed difference between words and nonwords might then disappear. At this time, we can only conclude that the covert naming of printed nonwords is not as efficient as that of words, which certainly agrees with previous results obtained in overt naming tasks.

No bias or facilitation of reaction time were obtained for words or nonwords in white noise. Whereas the signal-correlated noise masker forced subjects to extract speech-specific spectral information, essentially phonetic features, the white noise masker permitted use of simple auditory strategies: Any deviation from the random noise background, whether speechlike or not, could be used in the decision process. Given the very low S/N ratios used, the information on which the subjects' decisions were based probably was not speechlike enough to interact with phonetic top-down information.

In conclusion, we find support in our results for an interactive view of the processes of visual and auditory word perception, even though the early auditory processes of spectral feature extraction appear to be impermeable to top-down influences. The interaction of the visual and auditory word processing systems (in the direction we investigated, viz., from visual to auditory) seems to take place because of a rapid recoding of printed words into internal speech comparable in all respects to the output of the auditory phonetic module.

APPENDIX

Words and nonwords used in the Experiments.

PUBLIC	TEAMON
BODY	DEEMY
CLOSET	KETTER
BABY	BAXI
DOLLAR	DALIK
PICTURE	PIRTON
PAPER	PAMET
TEMPLE	TRISIN
PERSON	TILBER
PARENT	BEALTY
CARGO	PINOW
TABLE	TARNET
CANYON	TONKOR
CORNER	DORIT
TOTAL	PROSOR
CANVAS	BOONTER
DANGER	QUEMPLE
KITCHEN	BOTCHEN
PUPIL	PUNIL
DIMPLE	TUNY
PANIC	PAGER
PRISON	PROSOR
GARDEN	GASNET
PENCIL	CALVAS

ACKNOWLEDGMENT

This work was supported in part by National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories.

REFERENCES

Cole, R. A., & Jakimik, J. (1978). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-164). Hillsdale, NJ: Erlbaum.

Frauenfelder, U. H., & Tyler, L. K. (1987). The process of word recognition: An introduction. *Cognition, 25,* 1-20.

Frost, R., & Katz, L. (1989). Orthographic depth and the interaction of visual and auditory processing in word recognition. *Memory & Cognition, 17,* 302-310.

Glushko, R. J. (1979). The organization of activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 674-691.

Horii, Y., House, A. S., & Hughes, G. W. (1971). A masking noise with speech-envelope characteristics for studying intelligibility. *Journal of the Acoustical Society of America,* 49, 1849-1856.

Klatt, D. H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech.* Hillsdale, NJ: Erlbaum.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology.* New York: Wiley.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interaction during word recognition in continuous speech. *Cognitive Psychology, 10,* 29-63.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

McCusker, L. X., Hillinger, M. L., & Bias, R. G. (1981). Phonologic recoding and reading. *Psychological Bulletin, 89,* 217-245.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Repp, B. H., & Frost, R. (1988). Detectability of words and nonwords in two kinds of noise. *Journal of the Acoustical Society of America, 84,* 1929-1932.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110,* 474-494.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 43,* 1735-1736.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3-52). Hillsdale, NJ: Erlbaum.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *Journal of the Acoustical Society of America, 82,* 1152-1161.

Venezky, R. L. (1970). *The structure of English orthography.* The Hague: Mouton.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167,* 392-393

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America, 60,* 1381-1389.

FOOTNOTES

[*] *Journal of Memory and Language, 27,* 741-755 (1988).

[†] Now at the Department of Psychology, Hebrew University, Jerusalem, Israel.

[††] Also, University of Connecticut, Storrs.

[1] Given a maximum of 24 responses per subject and condition, values of 0.5 and 23.5 were substituted for response frequencies of 0 and 24, respectively, so as to obtain finite *d* and *b* indices.

[2] Note also that the S/N ratio for white noise was much lower than that for signal-correlated noise at a similar performance level. The ratios are not exactly comparable, however, because they have a different reference. S/N ratios would have been more similar if the white noise had been specified with reference to peak, rather than average, speech signal levels. (See Horii, House, & Hughes, 1971.)

[3] The claim that the generation of phonetic structure from print, is faster postlexically then prelexically does not imply that fast lexical access for words is achieved by a "visual route." Access to the lexicon may be achieved when sufficient phonologic information for determining a specific lexical entry has accumulated prelexically. However, such information may not be sufficient for the generation of a detailed phonetic structure.

Detectability of Words and Nonwords in Two Kinds of Noise*

Bruno H. Repp and Ram Frost[†]

Recent models of speech perception emphasize the possibility of interactions among different processing levels. There is evidence that the lexical status of an utterance (i.e., whether it is a meaningful word or not) may influence earlier stages of perceptual analysis. To test how far down such "top-down" influences might penetrate, we investigated whether there is a difference in detectability of words and nonwords masked by amplitude-modulated or unmodulated broadband noise. The results were negative, suggesting either that the stages of perceptual analysis engaged in the detection task are impermeable to lexical top-down effects, or that the lexical level was not sufficiently activated to have any facilitative effect on perception.

INTRODUCTION

In experimental psychology it is now common to conceptualize perceptual and cognitive processes in terms of parallel interactive networks. These models have also been applied to speech perception (e.g., Elman & McClelland, 1986; Samuel, 1986). Their architecture—a hierarchically layered system of interconnected informational units capable of being activated in varying degrees—permits rapid and parallel interchange of information between lower and higher levels in the hierarchy. These models thus assume that lower levels of perceptual analysis are not independent of higher levels: The input activates low-level units, which activate connected higher-level units, which in turn boost the activation of all lower-level units they are connected to, and so on until the system reaches some state of equilibrium.

An unconstrained version of such a model predicts interactions among all processing levels in the system. The task of experimental psychologists thus is to determine whether specific kinds of interactions do occur, and if they do not, to introduce constraints into the model. Thus, for example, there is evidence that influences of semantic context do not penetrate to earlier levels of analysis in speech perception (Connine & Clifton, 1987; Samuel, 1981). On the other hand, studies of phenomena such as phoneme restoration (Samuel, 1981) and the category boundary shift along word-nonword continua (Connine & Clifton, 1987) have suggested that the lexical status of an utterance (i.e., whether it is a meaningful word or not) may affect the extraction of phoneme information from an impoverished or ambiguous stimulus. It is not known, however, whether even earlier levels of perceptual analysis, such as the extraction of phonetic features from speech, can be affected by lexical status. In this study, therefore, we asked whether lexical status can affect the *detectability* of speech in noise.[1]

We were alerted to this possibility by our findings in a recent study (Frost et al., 1988), in which subjects were required to detect the presence of speech signals in amplitude-modulated (AM) noise while watching printed stimuli on a computer screen. In one experiment, we used disyllabic words; in another, disyllabic nonwords (i.e., pronounceable pseudowords) similar in structure. Subjects' detection performance was much worse with the nonwords, at exactly comparable speech-to-noise (S/N) ratios. Although this result could have been caused by the accompanying visual stimuli, it could also reflect a genuine difference in the auditory detectability of words and nonwords. Such a difference, if confirmed, would provide the strongest possible evidence for lexical "top-down" effects on speech perception, because a detection task taps into the earliest levels of perceptual processing. On the other hand, a demonstration that these early levels are immune to lexical influences would place a further constraint on interactive models of speech perception, and thereby would help reveal the internal architecture of the speech perception system.

The level of perceptual analysis employed in a speech detection task may depend on the kind of masking noise employed. In our earlier study, the masking noise always had the same amplitude envelope as the speech to be detected (cf. Horii et al., 1971). Subjects thus had to detect spectral evidence of speech against a background of appropriate speech envelope features (cf. Van Tasell et al., 1987). The spectral features in combination with the envelope features may have provided sufficient information for activation of a narrow range of lexical candidates, not enough for accurate identification (according to informal observations) but perhaps enough to generate a facilitative top-down flow to earlier perceptual stages. In unmodulated (UM) white noise, on the other hand, any kind of auditory evidence can be used for detection of speech, so that very little information enters the perceptual system from near-threshold stimuli. Nevertheless, we wondered whether effects of lexical status might be obtained even under those circumstances.

In the present experiment, then, we compared the detectability of structurally similar words and nonwords, presented randomly intermixed to the same subjects in either AM or UM noise. Apart from the issue of lexical effects, the comparison of speech detectability in AM and UM white noise was of some methodological interest (see Horii et al., 1971). In particular, we wondered whether the slope of the detectability function (percent correct detection as a function of S/N ratio in dB) would be shallower in AM than in UM noise (in view of the seemingly greater difficulty of detection in AM noise) or the reverse (in view of the matched amplitude envelopes of speech signal and AM noise, which may enable the speech signal to emerge suddenly as the S/N ratio is increased).

Methods

Subjects

Twelve paid undergraduate subjects participated. All were native speakers of English and reported having normal hearing.

Stimuli

We used the same words and nonwords as Frost et al. (1988). Each set comprised 24 disyllabic stimuli beginning with a stop consonant, each containing from four to six phonemes and being stressed on the first syllable. The word frequencies ranged from 0 to 438, with a median of 60, according to Kučera and Francis (1967). The nonwords were generated from a different set of disyllabic words by changing one or two phonemes, without violating the phonotactic rules of English. All stimuli were spoken by a young woman in a sound-insulated booth and were recorded using high-quality equipment. The utterances were digitized at a 20 kHz sampling rate, low-pass filtered at 9.6 kHz, and stored in separate computer files.

An AM noise masker was generated for each individual utterance by a computer program that reversed the polarity of sampling points with a probability of 0.5 (see Schroeder, 1968). Each masking noise thus had exactly the same amplitude envelope and overall level as its speech mate, but it had no spectral structure.[2] A UM masking noise was obtained for each individual stimulus by excerpting a segment of exactly the same duration from a longer file of white noise (sampled from a General Radio 1390-A random noise generator), applying 5-ms amplitude ramps to avoid abrupt onsets and offsets, and attenuating the noise until its average level matched that of the speech.

Speech-plus-noise stimuli were obtained by digitally adding the waveforms of a speech stimulus and each of its two matched noise maskers. Each masking noise thus began and ended with the speech. In the waveform-adding procedure, weights were applied to both digitized files to vary the S/N ratio while keeping the overall level of the added stimulus constant. Thus, the level of the speech decreased as that of the noise increased. Five S/N ratios, spaced 2 dB apart, were chosen on the basis of pilot data for each noise condition. They ranged from -18 to -10 dB for the modulated noise, and from -32 to -24 dB for the UM noise.[3] Since these ratios were all negative, changes in S/N ratio entailed primarily changes in the absolute level of the speech signal and only negligible changes in noise level.

Each of the two stimulus sets, words and nonwords, was divided randomly into two sets of 12, which were then combined to form two parallel stimulus sets, each containing 12 words and 12 nonwords. Each of these two sets was presented in each of the two noise conditions. The corresponding four stimulus sequences were recorded on audio tapes. Each of the tapes began with 36 familiarization trials, which were followed by 5 blocks of 24 experimental trials. The familiarization trials used a single word (*powder*), which was not contained in the experimental set. Signal-plus-noise and noise-only trials were presented in strictly alternating fashion, and no responses were required. Six groups of six familiarization trials represented S/N ratios of increasing difficulty, the first one being very easy, and subsequent ones being those used in the experimental trials. Each of the following experimental blocks contained one instance of each stimulus, at a fixed S/N ratio. Successive experimental blocks repeated the same stimuli in a different random order and simultaneously increased the S/N ratio by 2 dB. The task thus started with the most difficult condition and became progressively easier. The reason for this was to reduce practice effects, since the stimuli were the same in each block; however, there was only a slim chance of actually identifying any word or nonword at even the most favorable S/N ratio.

Procedure

Each subject listened to one AM noise tape and one UM noise tape containing different speech stimuli. Half the subjects listened to one pair of tapes and half to the other. Half the subjects received one noise condition first, and half the other. All subjects listened binaurally in a quiet room over TDH-39 earphones. The instructions emphasized that detection, not identification of the speech was required, and that it was a good idea to guess and use "yes" and "no" responses about equally often. The playback level was calibrated on a vol meter using white noise recorded at the beginning of the tape. The average levels of individual stimuli at the subjects' earphones ranged from 83 to 88 dB SPL.

Results

The results are shown in Figure 1. Performance (percent correct detections) increased steadily as S/N ratio increased, from near chance to between 85 and 90 percent correct in each noise condition. There was no difference between words and nonwords in either AM or UM noise. The slopes of the detectability functions were also similar in the two noise conditions.

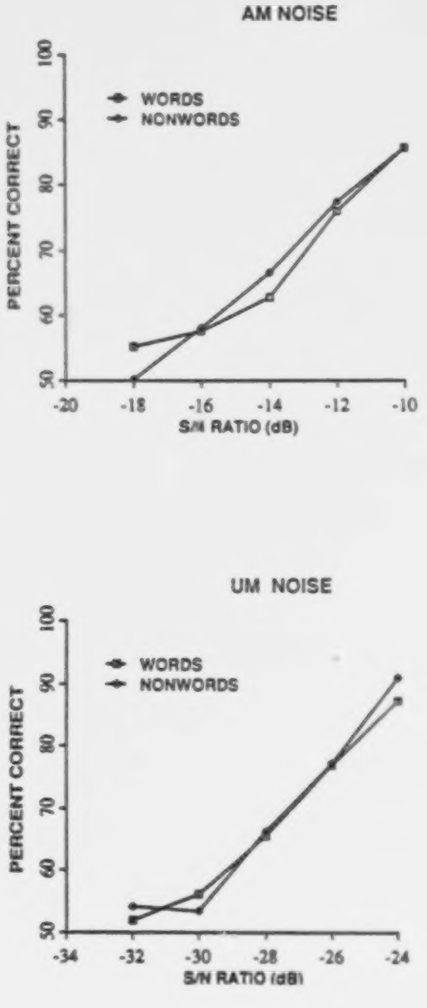


Figure 1. Percent correct detections as a function of S/N ratio for words and nonwords in two noise conditions.

Because the speechlike features of the AM noise may have caused a response bias to say "yes," the results were also analyzed in terms of separate indices of discriminability (sensitivity) and bias, following the procedures of Luce (1963). The two indices, referred to here as *d* and *b*, are defined as

$$d = (1/2)\ln[p(\text{yes} \mid S{+}N)p(\text{no} \mid N)/p(\text{yes} \mid N)p(\text{no} \mid S{+}N)]$$

and

$$b = (1/2)\ln[p(\text{yes} \mid S{+}N)p(\text{yes} \mid N)/p(\text{no} \mid S{+}N)p(\text{no} \mid N)]$$

where S+N and N stand for speech-plus-noise and noise alone, respectively. The discriminability index *d* assumes values in the same general range as the d' of signal detection theory, with zero representing chance performance. The bias index &b assumes positive values for a tendency to say "yes" and negative values for a tendency to say "no".[4]

The indices, averaged over subjects, are shown in Table I. The *d* indices confirm that performance was similar for words and nonwords, in both noise conditions. In a repeated-measures analysis of variance on these indices (with the factors noise type, S/N ratio, and lexical status), no effect except the obvious one of S/N ratio even remotely approached significance.

Table 1: Discriminability and bias indices as a function of lexical status, noise type, and S/N ratio.

	S/N ratio (dB)									
	AM noise					UM noise				
	-18	-16	-14	-12	-10	-32	-30	-28	-26	-24
Discriminability										
Words	0.26	0.33	0.56	1.46	2.01	0.12	0.28	0.71	1.43	2.05
Nonwords	-0.03	0.38	0.76	1.41	1.96	0.19	0.18	0.78	1.49	2.40
Bias										
Words	-0.40	-0.03	-0:08	-0.15	0.30	-0.52	-0.31	-0.14	-0.13	-0.11
Nonwords	-0.38	0.09	-0.03	0.34	0.20	-0.27	-0.04	-0.27	0.16	-0.04

The *b* indices show that subjects had a tendency to say "no" at the lowest S/N ratio, but as soon as performance rose above chance, there was no clear response bias in either direction. There were no differences in bias between words and nonwords or between noise types. This was confirmed in an analysis of variance on the *b* indices, which revealed only a significant effect of S/N ratio, $F(4,44) = 5.95$, $p = 0.006$.

DISCUSSION

Our results show that structurally similar words and nonwords are equally detectable in noise, even when the noise provides a background of appropriate speech envelope features. Thus it appears that lexical influences did not penetrate to the levels of perceptual analysis required for the detection task. The difference between word and nonword detectability in AM noise found in our previous study (Frost et al., 1988) must have had a different origin, probably related to the simultaneous visual presentation of words or nonwords. We conclude tentatively that the earliest stages of speech analysis are not permeable to lexical top-down effects, and that this needs to be taken into account in interactive models of speech perception.

Our conclusion is tentative because there are two other possible reasons for our negative results. First, stimulus information may have been too limited to lead to lexical activation that was sufficiently constrained to have a facilitative top-down effect. This was especially true in the UM noise condition, although in the AM noise condition, because of the combination of spectral and envelope features, there was a good chance of more focused lexical activation, especially at the higher S/N ratios. Even so, there was considerable uncertainty about the possible lexical choices, which contrasts with other tasks in which lexical top-down effects have been observed (Connine & Clifton, 1987; Samuel, 1981).

The other possible reason for our negative findings is that subjects in the detection task may have employed a purely auditory strategy. Even in AM noise, detection of the voice fundamental, for example, may have been sufficient. The subjects were not forced to detect phonetic features or phonemes as such. It is, of course, a matter of theoretical viewpoint whether or not early auditory analysis is considered to be part of the speech perception system. If the lowest level units are assumed to be phonetic features (as in the TRACE model of McClelland & Elman, 1986), then it may be argued that lexical top-down effects do not occur in a detection task because the speech perception system is not engaged.

In addition to the main finding of no difference between words and nonwords, our results reveal that the detectability functions in AM and UM noise are rather similar, despite the apparent difference in S/N ratios and the different task demands. Horii et al. (1971) examined the intelligibility ("articulation") functions for consonants and vowels embedded in a constant phonetic context, and also found similar functions in AM and (continuous) UM noise. It is interesting to note that, in their study, the recognition thresholds (70 percent correct) for both segment types were at a S/N ratio of approximately -13 dB in both types of noise, after adjusting the S/N ratios for UM noise by taking into account the different absolute levels of consonants and vowels. In the present study, the *detection* threshold for words and nonwords in AM noise was at about -13 dB. Even though the two studies are not directly comparable because of differences in materials and other factors, the comparison nevertheless suggests that there is only a small gap between the detection and recognition thresholds in AM noise. Since envelope features derived from the masking noise reduce the number of possible segmental alternatives (Van Tasell et al., 1987), any spectral features detected may be sufficient for homing in on the correct segment. In UM noise, on the other hand, there is a clear discrepancy between the detection and recognition thresholds, as was demonstrated long ago by Hawkins and Stevens (1950).

ACKNOWLEDGMENT

This research was supported by NICHD Grant HD-01994 to Haskins Laboratories.

REFERENCES

Broadbent, D. E. (1981). *Decision and stress.* London: Academic.

Connine, C. M., & Clifton, C. Jr. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 291-299.

Frost, R., Repp, B. H., & Katz, L. (1988). Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory and Language, 27,* 741-755.

Hawkins, J. E., & Stevens, S. S. (1950). The masking of pure tones and of speech by white noise. *Journal of the Acoustical Society of America, 22,* 6-13.

Horii, Y., House, A. S., & Hughes, G. W. (1971). A masking noise with speech-envelope characteristics for studying intelligibility. *Journal of the Acoustical Society of America, 49,* 1849-1856.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America, 29,* 296-305.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology.* New York: Wiley.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General,* 110, 474-494.

Samuel, A. G. (1986). The role of the lexicon in speech perception. In E. C. Schwab & H. C. Nusbaum (Eds.), *Patterns recognition by humans and machines,* (Vol. 1, pp. 89-111). New York: Academic Press.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America,* 43, 1735-1736.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *Journal of the Acoustical Society of America, 82,* 1152-1161.

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America, 60,* 1381-1389.

FOOTNOTES

[*] A shorter version has appeared in the *Journal of the Acoustical Society of America, 84,* 1929-1932 (1988).

[†] Now at the Department of Psychology, Hebrew University, Jerusalem, Israel.

[1] It is well known that familiar words are *identified* better than unfamiliar words in noise (e.g., Howes, 1957).Although this "word frequency effect" may reflect an influence of lexical representations on early perceptual processes, an interpretation in terms of response bias has generally been considered sufficient (see, e.g., Broadbent, 1971).

[2] After the random sample-reversal procedure, the noise had a flat spectrum, like white noise. However, because it had to be converted into sound at the same time as the speech, which had been input with high-frequency pre-emphasis, the noise underwent high-frequency de-emphasis at the output stage. Thus the AM noise, in contrast to the UM noise, actually had a sloping spectrum (about 6 dB per octave above 1000 Hz, and less below).

[3] The S/N ratios for the two types of noise were not directly comparable because the former were exact (i.e., they remained constant with changes in speech level within a stimulus), while the latter related to the average speech level only (i.e., they varied with local changes in speech level). They would have been more similar, had the UM S/N ratios been specified with respect to the peak level of the speech. (See Horii et al., 1971, for discussion of these issues.)

[4] Given a total of 12 responses per subject and condition, values of 0.5 and 11.5 were substituted for response frequencies of 0 and 12, respectively, so as to obtain finite *d* and *b* indices. See Wood (1976) for an earlier application of these indices in a speech perception task.

Discovering Phonetic Coherence in Acoustic Patterns*

Catherine T. Best,[†] Michael Studdert-Kennedy,[††] Sharon Manuel,[†††] and Judith Rubin-Spitz[††††]

Despite spectral and temporal discontinuities in the speech signal, listeners normally report coherent phonetic patterns corresponding to the phonemes of a language that they know. What is the basis for the internal coherence of phonetic segments? On one account, listeners achieve coherence by extracting and integrating discrete cues; on another, coherence arises automatically from general principles of auditory form perception; on a third, listeners perceive speech patterns as coherent because they are the acoustic consequences of coordinated articulatory gestures in a familiar language. We tested these accounts in three experiments by training listeners to hear a continuum of three-tone, modulated sine wave patterns, modeled after a minimal pair contrast between three-formant synthetic speech syllables, either as distorted speech signals carrying a phonetic contrast (Speech listeners), or as distorted musical chords carrying a nonspeech auditory contrast (Music listeners). While the Music listeners could neither integrate the sine wave patterns nor perceive their auditory coherence to arrive at consistent, categorical percepts, the Speech listeners judged them as speech almost as reliably as the synthetic syllables on which they were modeled. The outcome is consistent with the hypothesis that listeners perceive the phonetic coherence of a speech signal by recognizing acoustic patterns that reflect the coordinated articulatory gestures from which they arose.

INTRODUCTION

To master their native language, children must learn not only to listen, but to speak. In the speech signal, they must discover information that not only distinguishes among the words they hear, but also specifies how the words are to be spoken. This dual function of the speech signal has been largely disregarded in research on speech perception. Researchers have generally accepted the linguist's description of speech as a sequence of syllables or phonemes, compounded from "bundles of features," and have then looked in the signal for the "information-bearing elements" or "cues" that correspond to the linguist's abstract descriptors, without considering whether or how these cues might specify the articulatory gestures that give rise to them. The strategy has been successful to the extent that we now have detailed lists of cues—pitch contours, formant patterns, silent gaps, patches of band-limited noise, and so on—that may be mimicked by terminal analog synthesis to render intelligible speech.

Such synthesis typically proceeds, however, without appeal to general principles of either auditory or articulatory organization, or of their interrelationship. Even if an experimenter follows certain "rules for synthesis," the rules are rarely more than a summary of previous experimenters' prescriptions for copying spectrograms within the constraints of a particular synthesizing device (but see Mattingly, 1981). The criterion for a successful copy is simply a listener's judgment as to whether or not the synthesized pattern renders an acceptable phonetic form (i.e., an acceptable acoustic-articulatory pattern of sound) in the language under study.

What is the basis for listeners' phonetic percepts? What do they listen for in the signal? From the facts of speech synthesis, we might suppose that they listen for discrete acoustic cues. However, the notion of *cue extraction* poses a logical puzzle of definition, as noted by Bailey and Summerfield (1980). To establish that a particular piece of acoustic "stuff" deserves the status of a cue, researchers commonly use speech synthesis to set all other portions of the array at values that will ensure perceptual ambiguity. They then manipulate the potential cue, so that particular settings resolve the ambiguity. However, when the ambiguity is resolved—that is, when listeners consistently identify the pattern as an instance of a particular phonetic category—which is the cue? Is it the element that was manipulated or is it the context? The context without the cue is ambiguous, and the cue without its context is typically heard as nonspeech (e.g., Mattingly, Liberman, Syrdal, & Halwes, 1971). We have no grounds for preferring one to the other as an effective or necessary component of the pattern.

If neither cue nor context (itself composed of an indefinite number of other cues) can independently and unambiguously specify the speech sounds we hear, the functional unit of speech perception must be the entire acoustic pattern that the acoustic cues compose. What is this pattern; and why do the diverse "cues" that compose it cohere perceptually?[1]

According to one account listeners extract discrete cues, but judge them only in relation to each other, so that the phonetic segment is a result of their perceptual integration (e.g., Cutting, 1976; Jusczyk, Smith, & Murphy, 1981; Pastore, 1981; Schouten, 1980). The reason why isolated cues are often heard as nonspeech is that perceptual categorization depends on the relations among cues, and these relations are destroyed when a cue is removed from context. A variant of this view treats the supposed cues as independent "features" to which listeners assign weights on the basis of their representation in the signal. Listeners then sum or multiply the weights and compare the integrated outcome with a stored "prototype" to arrive at a probabilistic estimate of the percept (Oden & Massaro, 1978). We will refer to the mechanism proposed by these accounts as *cue integration.*

According to another account, cues have no functional role in determining the sound pattern of speech. Rather, the pattern coheres according to Gestalt principles analogous to those in visual form perception, such as proximity, similarity, good continuation and closure (Bregman, 1981). Thus, the melodic coherence of vowel sequences, essential to prosody, may be maintained across consonantal constrictions by the smooth contour of their fundamental frequencies (Bregman, 1981). The harmonics of a vowel formant may cohere by virtue of temporal proximity, that is, of their simultaneous onsets and offsets (Darwin, 1984). Temporal proximity may also account for coherence of the spectrally diverse cues to voicing in consonant-vowel (CV) syllables, discussed below. Good continuation and spectral similarity may be at work in a CV syllable when a stop consonant release burst effectively conveys information about place of articulation only if it is spectrally continuous with the following formant transition

(Dorman, Studdert-Kennedy, & Raphael, 1977). Finally, formant transitions may perform not only segmental functions, but also syllabic functions, by eliminating from the signal abrupt discontinuities that might excite an unwanted increase in neural firing (Delgutte, 1982). The transitions would thus assure syllabic coherence and, incidentally, correct perception of the temporal order of syllabic components in rapid speech (Cole & Scott, 1974; Dorman, Cutting & Raphael, 1975). If this account is correct, phonetic forms emerge from the signal by virtue of their *auditory coherence*. Notice that this account, unlike that based on cues, has nothing to say about the units of linguistic information that the speech signal conveys. Principles of auditory coherence are presumed to apply not only to phonetic segments, but to every other unit of linguistic analysis, from the feature to the prosodic contour.

A final account invokes a principle that we will call *phonetic coherence*. The basis for perceptual coherence, according to this account, is said to be the coordinated pattern of articulatory gestures that produced the signal. The principle is implicit in the well-known explanation offered over 20 years ago by Abramson and Lisker (1965) for the spectral and temporal diversity of covarying cues to voicing distinctions in many languages: release burst intensity, degree of aspiration and first formant (F1) onset frequency. They proposed that all these cues arise from the relative timing of laryngeal and supralaryngeal gestures in stop-vowel syllables:

> "Laryngeal vibration provides the periodic or quasi-periodic carrier that we call voicing. Voicing yields harmonic excitation of a low frequency band during closure, and of the full formant pattern after release of the stop. Should the onset of voicing be delayed until some time after the release, however, there will be an interval between release and voicing onset when the relatively unimpeded air rushing through the glottis will provide the turbulent excitation of a voiceless carrier, commonly called aspiration. This aspiration is accompanied by considerable attenuation of the first formant, an effect presumably to be ascribed to the presence of the tracheal tube below the open glottis. Finally, the intensity of the burst, that is, the transient shock excitation of the oral cavity upon release of the stop, may vary depending on the pressures developed behind the stop closure where such pressures will in turn be affected by the phasing of laryngeal closure. Thus it seems reasonable to us to suppose that all these acoustic features [cues], despite their physical dissimilarities, can be ascribed ultimately to actions of the laryngeal mechanisms." (Abramson & Lisker, 1965, pp. 1-2).

Variations in voice onset time underlie voicing distinctions in many, if not all, languages, and this elegant account of the articulatory origin of the diverse cues to voicing has been widely accepted. What is important here, however, is the general principle that the model proposes: the speech signal coheres not because of (perhaps even in spite of) its auditory properties, but because coordinated patterns of gesture (i.e., of phonetically functional articulatory actions, such as lip closure, velum lowering, tongue raising, etc.), give rise to coordinated patterns of spectral and temporal change. By adopting the articulatory gesture, and its acoustic correlates, as its linguistic primitive, this account proposes an objectively observable unit common to both production and perception. Gestures thus form both the patterns of information that listeners listen for in synthetic speech, and the patterns that children must discover in natural speech if they are to learn how to talk. Thus, the *phonetic coherence* account is the only account discussed here to offer a direct, concrete basis for the perception-production link, and hence, for imitation. This gestural account is compatible with any level of abstract linguistic unit from the phoneme (Fowler & Smith, 1986; Studdert-Kennedy, 1987) to the word (Browman & Goldstein, 1986).

To summarize, each of these accounts offers a view, implicit or explicit, of (1) the information (i.e., the linguistic structure) that a speaker encodes in the signal, and (2) the mechanism by which a listener recovers that structure. Both the *simple cue extraction* and the *cue integration* accounts propose that the information is a collection, or sequence, of abstract linguistic elements—features, phonemes, or perhaps syllables—and that the recovery mechanism entails the simple extraction, or the extraction and integration, of discrete "cues" to those elements. The *auditory coherence* account, the least linguistically oriented, is neutral on the nature of the linguistic information, but proposes that listeners perceive the sound patterns of speech (whatever they may be) according to general principles of auditory form perception. Finally, the *phonetic coherence* account proposes that the linguistic structure of the signal is articulatory, a pattern of gestures, and that listeners recover this structure because it is implicit in the acoustic signal to which it gives rise. Whether the articulatory gestures are grouped and segregated so as to specify abstract units at an intermediate phonological level (phonemes, syllables) or only at the level of lexical items (morphemes, words) is a separate issue, not considered here.

The following three experiments were designed to test these accounts of speech perception. First, they bring further experimental evidence to bear on the arguments presented above concerning the role of cues in speech perception: They ask whether listeners can better learn to identify, and discriminate between, contrasting acoustic patterns by focusing attention on a discrete acoustic cue, or by focusing on the entire acoustic pattern of which the cue is a part. Second, if attention to the entire pattern yields superior performance, the experiments are so designed that we can ask further whether the contrasting patterns emerge according to principles of cue integration or auditory form perception, or from listeners' directing attention to their potential phonetic coherence.

We compared the perceptual effects of an attentional focus on a phonetic contrast, /r/ vs. /l/, with the effects of attention to a discrete acoustic cue signaling that contrast, both in and out of context. Our stimulus materials were a continuum of sine wave speech syllables. Sine wave speech can be heard either as distorted, though recognizable, speech, or as sounds unrelated to speech (e.g., distorted musical chords or bird-like chirps) (Best, Morrongiello, & Robson, 1981; Remez, Rubin, Pisoni, & Carrell, 1980). This dissociation allowed us to compare perceptual responses to the same signal under nonspeech (auditory) and speech (phonetic) modes of attention.

EXPERIMENT 1

The first experiment investigated which of the views outlined above best accounts for the discrete perceptual categories in a minimal pair speech contrast: simple cue extraction, cue integration, auditory coherence or phonetic coherence.

To test these possibilities, we developed a sine wave syllable continuum based on the time-varying formant frequencies characteristic of American English /ra/ vs. /la/. The members of this continuum differed only in the direction and rate of the third formant (F3) transition, which varied systematically in approximately equal steps. Two additional series were developed for control comparisons: synthetic full-formant versions of the syllable continuum and frequency-modulated single tones corresponding to the F3 elements of the sine wave syllable series. Listeners participated in one of two conditions: speech bias or music bias. Pre-test instructions and perceptual training tasks were designed to focus the speech listeners' attention on hearing the sine wave syllables as distorted versions of /ra/ and /la/, the music listeners' attention on hearing them as carrying a binary contrast (steady vs. rising) on the transition of the F3 tone, the highest tone in a distorted three-tone musical chord.

If the speech categories depend on extraction of a single cue (F3 transition), both groups of listeners should perform identically in categorizing both the sine wave continuum and the full-formant continuum: the 50 crossover points on their identification functions (i.e., their category boundaries) and the slopes of these functions should be the same. Moreover, both groups should categorize the isolated single tones (F3) exactly as they categorize the sine wave and full-formant syllables.

If the speech categories depend on the extraction and integration of acoustic cues, we would expect that music listeners, trained to attend to the sine wave F3 transition, should be able to extract and integrate that transition no less consistently than speech listeners. They may assign a different weight to the transition in the overall structure, and thus discover a somewhat different perceptual organization with a different category boundary from that of the speech listeners, but the consistency (i.e., the slope) of the sine wave categorization functions should be roughly the same for the two groups.

Similarly, if the speech categories reflect the operation of general Gestalt principles of auditory form perception, music listeners should be able to discover at least some consistent pattern (either the same as or different from that of the speech listeners) in the sine wave stimuli. Thus, on the auditory coherence account, we would again expect the slopes, if not the category boundaries, to be essentially the same for both groups.

Finally, if the speech categories depend on phonetic coherence, the two groups should differ in their judgments of the sine wave syllables. The speech listeners should recognize the phonetic coherence of the sine wave syllables and should categorize them as they categorize the full-formant versions, with perhaps somewhat shallower slopes (i.e., greater response variability) for the sine waves, due to the phonetically impoverished and unfamiliar patterns of the sine wave "dialect." By contrast, the Music listeners (unable to suppress the perceptual influence of the lower tones, or perhaps to reject the automatically given auditory coherence of the sine wave patterns) should have difficulty in separating the F3 tone perceptually from its chordlike frame, and therefore in categorizing the sine wave syllables on the basis of the F3 binary contrast. Their difficulties should be evident in flat sine wave syllable slopes, or at least in significantly shallower slopes for these patterns than the speech listeners show. Thus, the main test of the competing hypotheses lies in the effects of instruction condition on the slopes of the sine wave syllable functions.

Method

Subjects

Twenty-four subjects were tested in a between-groups design, with 12 subjects each in the speech bias group (3 males, 9 females) and in the music bias group (7 males, 5 females) (Experiment 1a). In addition, 5 of the music subjects returned for a second session under speech bias conditions, permitting a partial within-group test (Experiment 1b). It was not possible to conduct a full within-group study, with condition orders counterbalanced, since our own and other researchers' experience with sine wave speech indicates that once subjects have perceived the stimuli as speech, it is extremely uncommon for them to be able to revert to hearing them again as nonspeech.

All subjects were young adults with normal hearing, and with negative personal and family history of language and speech disorders. Each was paid $4 per test session. Based on their answers to posttest questionnaires (see *Procedure*), one female subject was eliminated from the speech group for failure to hear any of the sine wave syllables as speechlike, and one female was eliminated from the music group because she had heard the sine wave syllables as "sounding like r," thus reducing the n of each group to 11. This was necessary because evaluation of the hypotheses depended on consistent group differences in hearing the sine wave syllables as either speech or nonspeech.

Stimuli

The full-formant /ra/-/la/ series was developed first, using the OVE-IIIc serial resonance synthesizer. The continuum contained 10 items, differing from each other only in the onset frequency, and in the duration of the initial steady-state portion, of the F3 transition (see Figure 1). These F3 properties were varied in nearly equal steps (slightly constrained by the step-size limitations of the synthesizer). Each stimulus was 330 ms in duration. The series was designed to be biased toward perception of more /la/ than /ra/ tokens. This was done so that the phonetic category boundary should fall neither at the perceived shift from steady to rising F3 transitions nor at the continuum midpoint, since these physical properties were two likely foci for a psychoacoustically-based category distinction.

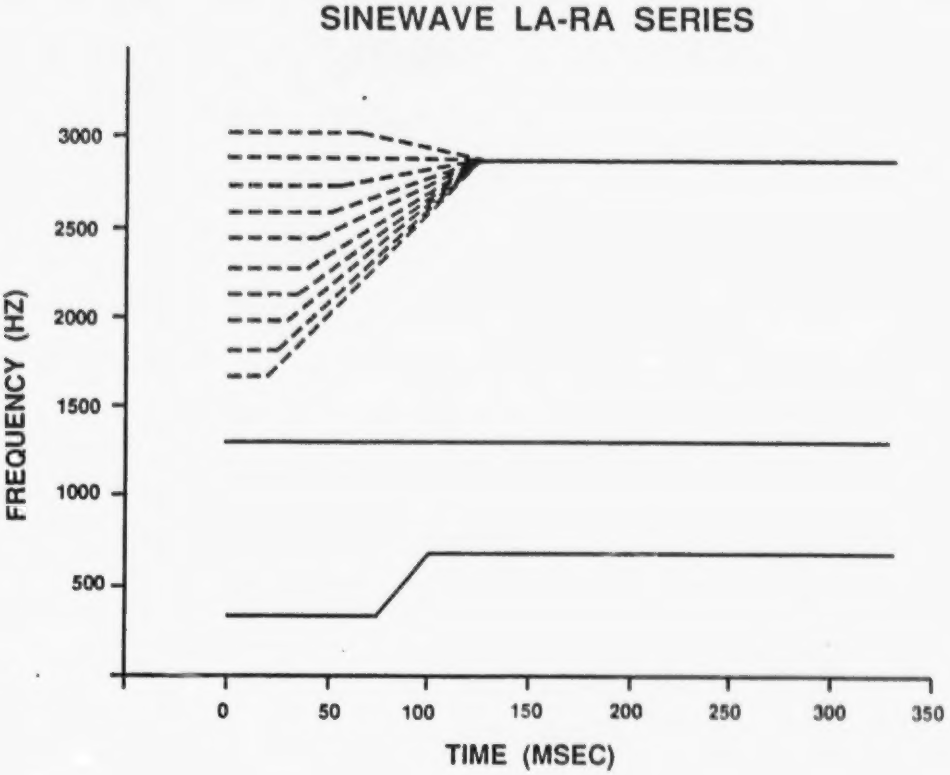


Figure 1. Schematic diagram of the center frequencies of the three formants in each of the sinewave syllables, and each of the full-formant syllables, in the 10-item stimulus continua.

The /l/ biasing was accomplished by using a rapid F1 transition, a long-duration steady-state at the onset of F1, and a steady-state F2, in the portion of the stimuli that was kept constant throughout the series, and by including one stimulus (the first on the continuum) with a slightly falling F3 transition. These formant characteristics are associated with natural tokens of /l/; natural /r/ typically has short steady-states at the onsets of the formant transitions, a relatively slow F1 transition, a slightly rising F2 transition, and a clearly rising F3 transition (MacKain, Best, & Strange, 1981). In all stimuli F0 began at 119 Hz and fell steadily to 100 Hz by the end of the stimulus. In the constant portion of the stimuli, F1 onset began at 349 Hz, remained there for 75 ms, then rose linearly to 673 Hz by 100 ms , where it remained to the end of the syllable. The steady-state frequency of F2 was 1297 Hz. The constant portion of F3 was a steady-state 2870 Hz in the final part of the stimulus, beginning at 125 ms into the syllable. The F3 onset frequencies for the 10 stimulus items were 3019 (at the /la/ end of the continuum), 2870 (a flat F3), 2729, 2576, 2431, 2278, 2119, 1972, 1821, and 1670 Hz (at the /ra/ end of the continuum). In the first stimulus (/la/), the steady-state onset ended and the F3 transition began at 65 ms into the stimulus. In the tenth stimulus (/ra/), this breakpoint occurred at 20 ms into the stimulus. The temporal position of the breakpoint was varied systematically in 5 ms steps for the intervening stimuli.

The sine wave syllable continuum was generated with a multiple sine wave synthesizer program developed for the PDP-11/45 computer at Haskins Laboratories. The frequency characteristics of the sine wave syllables mimicked those from the full-formant continuum, except that each formant was now represented as a single, time-varying tone rather than as the wider band of harmonics found in the formants of natural and synthetic speech. In the sine wave syllables, there was no tone to represent the original F0 contour (see Figure 1). The isolated F3 tone continuum was made up of the F3 tones from the sine wave syllable continuum, presented without the tones corresponding to F1 and F2.

Four additional stimulus series were developed for the perceptual training sequences that were presented in each condition before the categorization test for the sine wave syllables. These series were designed to focus the listener's attention either on the phonetic properties of the sine wave syllables, or on their properties as three-note chords differing only in the onset characteristics of the highest note. There were two speech bias training series, one based on the endpoint /ra/ stimulus of the sine wave syllable continuum, and the other based on the sine wave /la/ syllable with the flat F3 (the second stimulus in the continuum). Each speech training series contained 11 stimuli that provided a gradual, stepwise change from the full-formant syllable to the corresponding sine wave syllable. The first item of each series was the pure full-formant version of the syllable; the last item was the pure sine wave version of the syllable. The nine intervening stimuli were produced by mixing the exactly synchronized, matching sine wave and full-formant syllables in inversely varied proportions, i.e., the relative amplitude of the full-formant stimulus was reduced in equal steps, while the amplitude of the sine wave syllable was correspondingly increased. The two music bias training series also consisted of 11 items each, but the transformations progressed from the isolated endpoint F3 tone at the /ra/ end of the continuum to the corresponding endpoint sine wave syllable, and from the isolated flat F3 tone at the /la/ end of the continuum to the corresponding sine wave syllable.

Procedure

The subjects were tested in groups of 2 to 5 in a sound-attenuated experimental room. The stimuli were presented to them at a comfortable listening level (75 dB SPL) over TDH-39 headsets.

All subjects first completed the experimental task, consisting of the appropriate perceptual training sequence for the condition randomly assigned to their group (speech bias or music bias), followed by the categorization test with the sine wave syllables. This task was administered first so that the subjects' performance with the sine wave syllables could not be influenced by exposure to the full-formant and isolated F3 sine wave continua. The speech subjects were instructed that they would be tested on their ability to categorize computer-distorted versions of the syllables /la/ and /ra/, whereas the music subjects were instructed that they would be asked to categorize computer-distorted chords according to whether or not there was a rising frequency glide at the onset of the highest tone in the chords.

The subjects were then told that they would first receive some perceptual training to aid in focusing their attention on the identities of the distorted syllables (speech group), or on the steady-state vs. upgliding properties of the highest notes in the distorted chords (music group). Each training sequence proceeded in five steps. The speech subjects first heard the pair of full-formant clear-case stimuli (/la/ and /ra/) repeated five times, whereas the music subjects heard five repetitions of the clear-case F3 tones (flat and rising onsets), with 1-sec interstimulus intervals (ISIs) and 3-sec intertrial intervals (ITIs). Next, the subjects completed a 10-item practice test to categorize a randomized sequence of these clear-case syllables or F3 tones, presented individually with 3-sec ISIs, by entering their choices on an answer sheet. Third, they listened to, but did not explicitly categorize, the gradual 11-step transformation, beginning with the pair of full-formant syllables or the pair of F3 tones, and ending with the clear-case pair of computer-distorted sine wave stimuli. This transformation was then played in reverse order. The forward and reverse transformation was played three times, with ISIs of 3-sec, and interblock intervals (IBIs) of 6-sec. Fourth, subjects heard the pair of clear-case sine wave stimuli presented five times with 1-sec ISIs and 3-sec ITIs. Finally, they were given a randomized 20-item practice sequence of the clear-case sine wave stimuli presented one at a time with 3-sec ISIs, and they wrote their choices on answer sheets. For all practice trials, the correct answers were printed on the answer sheets but were covered by a strip of paper; the subjects uncovered each correct answer only after writing down their own response.

Subjects in both conditions then took a categorization test with the complete series of sine wave syllables. The test contained 20 blocks of the 10 items in the sine wave syllable continuum, randomized within each block. The stimuli were presented individually, with 3-sec ISIs and 6-sec IBIs. Subjects in the speech group circled "la" or "ra" on their answer sheets to indicate the category identity of each item in the test. Music subjects circled "steady" or "upglide" to indicate whether the highest tone in the distorted chord had a flat frequency trajectory or a rising glissando at the onset.

After they had finished the sine wave syllable test, subjects in both groups completed categorization tests with the two control series, the full-formant syllables and the isolated F3 tones. Each control test contained 20 randomized blocks of the 10 items in a given series, with 3- sec ISIs and 6-sec IBIs as before. On the full-formant syllable test, all subjects circled "la" or "ra" to indicate category assignments, and on the F3 tone test they circled "steady" or "upglide."

At the end of the test session, each subject answered a questionnaire about what the sine wave syllables had sounded like to them, whether they had been able to maintain the perceptual focus intended by their group's instructions, and whether they had made any judgments on the basis of the opposing group's perceptual set, i.e., whether the speech listeners had categorized the stimuli on the basis of musical (or nonspeech) properties, and whether the music listeners had heard any as syllables.

Results

Experiment 1a: Between-groups comparison

The categorization data were tabulated, for each subject on each continuum, as the percentage of times that each item was categorized as "la" in the full-formant syllable test and in the speech condition of the sine wave syllable test, or categorized as "steady" in the F3 tone test and in the music condition of the sine wave syllable test.

Figure 2 displays the averaged results for the two groups. The category boundaries (50% crossovers) fall at somewhat different points on the three continua (highest for the sine wave syllables, lowest for the F3 tones), but are essentially the same for the two groups. The slopes of the functions clearly differ across continua (steepest for the full-formant syllables, shallowest for the sine wave syllables), and are roughly the same for the two groups on the F3 tones and full-formant syllables. The groups differ markedly on the slopes of their sine wave syllable functions, with the slope for the music listeners being the shallower.

To test the significance of these effects, the data of each subject were first submitted to a probit analysis (Finney, 1971), to determine the mean (category boundary) and slope (reciprocal of the standard deviation) of the best-fitting ogive curve by the method of least squares.[2] Table 1 lists the mean category boundaries and slopes, together with their standard errors, for each group on each continuum. High slope values indicate steep slopes; low values indicate shallow slopes. Note that the computed mean category boundaries and slopes, based on individual probit analyses, are not expected to correspond exactly with those read from the group functions of Figure 2 (e.g., the relatively steep mean slope value of 2.30, listed in Table 1 for the speech listeners on the full formant syllables, was due to very steep slopes given by 3 out of the 11 listeners, but is not apparent in the group function of Figure 2). On the sine wave syllables, the music listeners were less consistent than the speech listeners in their category boundaries (higher standard error), but more consistent in their (low) slope values (lower standard error).

TABLE 1. Experiment 1a: Mean category boundaries and slopes, determined from individual probit analyses on stimulus items 2 through 10, for speech and music listeners.

	Category Boundaries			Slopes		
	Sine wave	F3	Full-formant	Sine wave	F3	Full-formant
			Speech Listeners (n=11)			
M	6.24	5.13	5.56	0.80	1.03	2.30
SE	(0.17)	(0.26)	(0.27)	(0.15)	(0.18)	(0.54)
			Music Listeners (n=11)			
M	6.69	5.18	5.67	0.14	0.96	1.70
SE	(1.16)	(0.24)	(0.20)	(0.05)	(0.10)	(0.35)

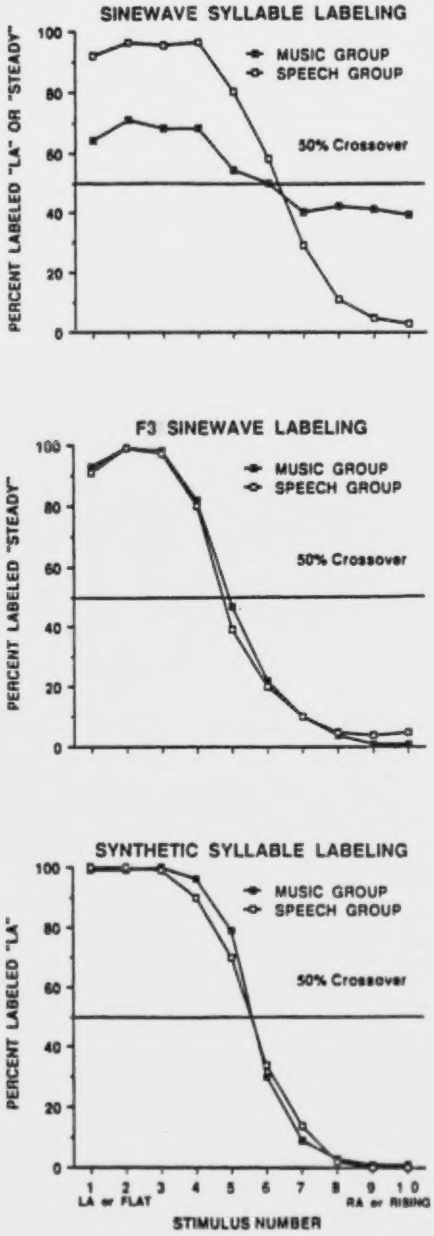


Figure 2. Labeling functions for the speech and music listeners on each of the three stimulus continua in Experiment 1a.

Two-factor (instruction group × stimulus continuum) analyses of variance were then carried out on the category boundaries and on the slopes. In these analyses, instruction group was a between-groups variable, stimulus continuum a within-groups variable. The category boundary analysis yielded no significant effects and no significant interaction. The slope analysis yielded a significant effect of stimulus continuum, *F*(2,

40) = 15.21, $p < .001$, an effect just short of significance for instruction condition, $F(1, 20) = 3.59$, $p < .07$, and no significant interaction, $F(2, 40) = 0.67$, $p > .10$. Scheffé tests of the stimulus continuum effect showed that slopes were significantly steeper for the full-formant syllables than for the F3 tones, $F(1,20) = 8.90$, $p < .05$, and steeper for the F3 tones, than for the sine wave syllables, $F(1,20) = 14.19$, $p < .02$.

The lack of an interaction in the slope analysis was unexpected, given the functions illustrated in Figure 2. Presumably, the fact that the music group had a shallower mean slope than the speech group on all three continua gave rise to the marginal effect of instruction condition, but the interaction failed to reach significance due to the large component contributed to the error variance by performance on the full formant syllables (see Table 1). Nonetheless, since a test of the slope difference between instruction groups on the sinewave syllables was a key to distinguishing among the competing hypotheses, we carried out a planned comparison by means of a simple t-test on these data (for which the error variance was relatively low: See Table 1). The result was highly significant, $t(20) = 4.18$, $p < .0005$.

Finally, as a test for category formation on the sine wave syllable continuum, we carried out for each group a one-factor analysis of variance with repeated measures on stimulus items. For the Speech listeners there was a significant effect of stimulus item, $F(8,80) = 110.18$, $p < .0001$; Scheffé tests between all possible pairs yielded significant differences between all items in the group 2 through 5 and all items in the group 7 through 10 ($p < .05$), but none between items within these groups, indicating the presence of two distinct categories. For the music group there was a significant effect of stimulus item, $F(8,80) = 5.88$, $p < .0001$; but Scheffé tests between all possible item pairs yielded significant differences only between item 2 and items 7 and 10, indicating no consistent categorization.

Experiment lb: Within-groups comparison

The data for the 5 subjects who served in both instruction conditions were treated in the same way as the data of Experiment 1a.[3] Figure 3 displays the group functions, and Table 2 lists the mean category boundaries and slopes, with their standard errors. The general pattern of results is similar to that of the between groups comparison. On the sine wave syllables the Music condition yields a higher standard error than the Speech condition for the category boundaries, a lower standard error for the slopes.

A two-factor analysis of variance with repeated measures on both factors (instruction condition and stimulus continuum) was carried out on the category boundaries and on the slopes. The category boundary analysis yielded a significant effect of stimulus continuum, $F(2,8) = 9.70$, $p < .01$, but no effect of instruction condition and no significant interaction. None of the three possible pair-by-pair category boundary comparisons between stimulus continua was significant on *post hoc* Scheffé tests. The slope analysis yielded a significant effect of stimulus condition, $F(2,8) = 6.08$ $p < .02$, but no effect of instruction condition and no significant interaction. None of three pair-by-pair slope comparisons between stimulus continua was significant by Scheffé tests. Again, despite the lack of an interaction, we carried out the planned *t* test (for matched pairs) to compare the slopes of the sine wave syllable functions between the two instruction conditions. The result was highly significant, $t(4) = 6.02$, $p < .004$.

Tests for category formation on the sine wave syllable continuum yielded a significant effect of stimulus item for the speech condition, $F(8,32)24.27$, p.0001, with items 2 through 5 and 7 through 10 again falling into distinct categories, according to Scheffé tests ($p < .05$ for all between-category comparisons, but not significant for any within-category comparisons). For the music condition there was a significant effect of stimulus items, but none of the Scheffé tests between item pairs was significant, indicating the absence of clearcut categories.

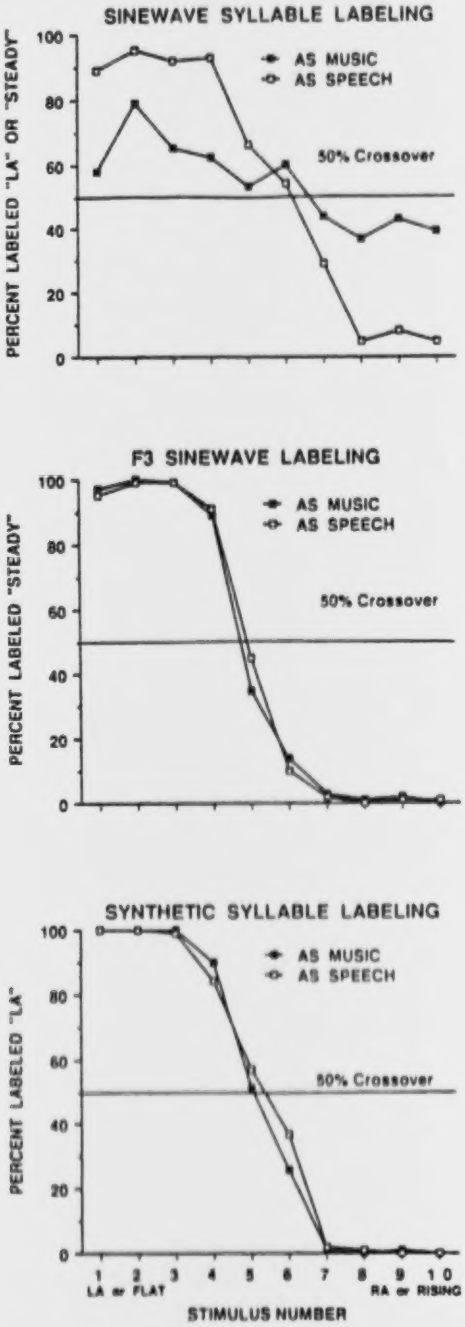


Figure 3. Labeling functions for the speech and music listeners on each of the three stimulus continua in Experiment 1b.

TABLE 2. Experiment 1b: Mean category boundaries and slopes, determined from individual probit analyses on stimulus items 2 through 10, for the 5 listeners who served in both speech and music instruction conditions.

	Category Boundaries			Slopes		
	Sine wave	F3	Full-formant	Sine wave	F3	Full-formant
			Speech Listeners (n=11)			
M	5.97	5.00	5.31	0.71	1.20	1.54
SE	(0.43)	(0.11)	(0.44)	(0.14)	(0.17)	(0.20)
			Music Listeners (n=11)			
M	7.15	4.75	5.48	0.14	1.34	1.79
SE	(0.68)	(0.19)	(0.21)	(0.07)	(0.20)	(0.64)

Discussion

The significant effect of stimulus continuum in the within-group comparison of category boundaries disconfirmed the prediction of the simple cue extraction hypothesis that boundaries would be identical on the three continua. There were no grounds in either experiment for rejecting the null hypothesis of equal category boundaries on the sine wave and full-formant continua.

The most decisive results came from the slope analyses. The two groups performed identically on the two control continua (F3 tones and full-formant syllables), but differed on the sine wave syllable continuum. Here, the significantly shallower slopes for the music than for the speech listeners, on both between-groups and within-group comparisons, suggest that attention to phonetic properties of the syllables facilitated categorization, whereas attention to purely auditory properties hindered it. This outcome is predicted only by the phonetic coherence hypothesis.

Nonetheless, as the stimulus item analysis indicated, the identification function of the music listeners on the sine wave syllables was not flat. Although they gave no evidence of reliable category formation, the music listeners' functions sloped in the same direction as that of the speech listeners in both experiments. We were therefore concerned that the music listeners might have been influenced by factors other than attentional mode. Specifically, the music group might have been disadvantaged by lack of practice with the crucial acoustic features of the category exemplars (i.e., steady vs. rising F3 tones) before the categorization test on the sine wave syllables. In addition, the words "steady" vs. "upglide" may have been more arbitrary as labels for the sine wave syllable endpoints in the music condition than were the "la" and "ra" labels in the speech condition.

EXPERIMENT 2

To meet the foregoing objections, we performed a second experiment in which we had each group first complete the control test that would constitute practice for the sine wave syllable categorizations (the F3 categorization test for the music listeners, the full-formant syllable test for the speech listeners) before completing the sine wave syllable test itself. We also provided the music listeners with nonverbal symbols of the endpoint F3 trajectories (—— vs. ╱——) to use as category identifiers in the sine wave syllable test rather than the perhaps arbitrary verbal labels used in Experiment 1.

Method

Subjects

Thirty young adults were tested. Of these, 13 were tested in the music bias condition (six males, seven females) and 17 were tested in the speech bias condition (four males, 13 females). On the basis of their answers on the posttest questionnaires (see Experiment 1), six subjects were eliminated from the analyses, leaving a total of 12 subjects in each group. One female was withdrawn from the music group because she began to hear words or names in the sine wave syllables, while five subjects (one male, four females) were withdrawn from the speech group for failing to hear the sine wave speech patterns as syllables. It may be of interest that of the latter participants, one was not a native speaker of English, and another was dyslexic. All remaining subjects lacked any personal or familial history of language and speech problems, were monolingual English speakers, and had normal hearing. Each received $4 for their participation in the test session.

Stimuli

The stimuli designed for Experiment 1 were used again in this experiment.

Procedure

The procedures were identical to those described in Experiment 1, except in the following respects. The music bias group used nonverbal labels (—— vs. ╱——) rather than the words "steady" and "upglide" to identify the items in the sine wave syllable and isolated F3 tone continua, and they completed the F3 categorization test before the training sequence and test with the sine wave syllable continuum. They did not take a categorization test with the full-formant syllables. The speech bias group, on the other hand, completed the categorization test with full-formant syllables before the training and test with the sine wave syllables. They again used "la" and "ra" as their category labels. They did not take a test with the isolated F3 tones.

Results

The data were treated as in the previous experiment. Figure 4 displays the group functions, and Table 3 lists the mean category boundaries and slopes, with their standard errors. For the speech listeners the category boundary is somewhat higher, and the slope shallower, on the sine wave syllables than on the full-formant syllables and the F3 tones. For music listeners the projected category boundary falls outside the continuum for the sine wave syllables, because most of the stimuli were labeled as " —— " (steady) across the whole series; their F3 category boundary falls well below the continuum midpoint. The sine wave syllable slope for the music listeners is close to zero, very much shallower than on the F3 tone series. The two groups differ strikingly in both category boundary and slope on the sine wave syllable continuum. Once again, the pattern of standard errors indicates that the music listeners were highly variable in their sine wave syllable boundaries, but highly consistent in their low slope values.

To test the significance of the category boundary and slope variations, we carried out *t* tests for correlated samples on the sine wave syllable and control continua for each group, and *t* tests for independent samples, comparing the groups on the sine wave syllable continuum. For the speech listeners, category boundaries on sine wave and full formant syllables did not differ, but slopes differed significantly, being shallower for the sine wave syllables, $t(11) = 3.45$, $p < .05$. For the music listeners, despite the large difference in means, category boundaries did not differ significantly on sine wave syllables and F3 tones, presumably due to the very high variability of their data. However, their slopes on the two continua differed significantly, being shallower for the sine wave syllables, $t(11) = 3.84$ $p < .0034$. The two groups did not differ in their category boundaries on the sine wave syllable continuum (presumably again due to the high variability of the music listeners), but they did differ significantly in their slopes, $t(22) = 5.62$ $p < .0001$.

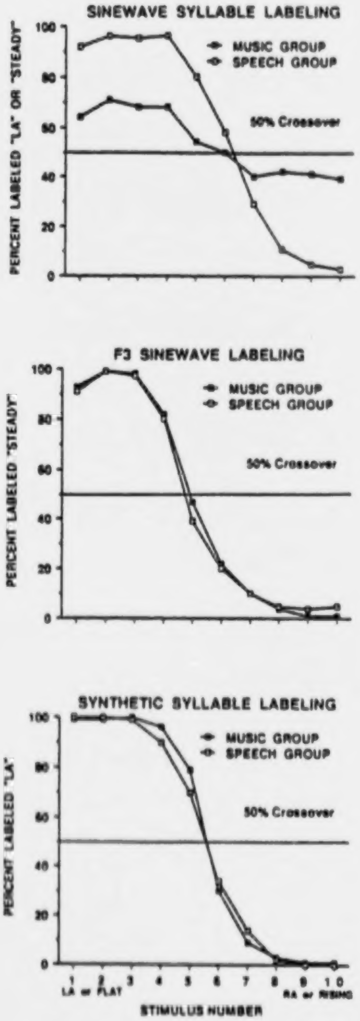


Figure 4. Labeling functions for the speech and music listeners on each of the three stimulus continua in Experiment 2.

TABLE 3. Experiment 2: Mean category boundaries and slopes, determined from individual probit analyses on stimulus items 2 through 10, for speech and music listeners.

	Category Boundaries			Slopes		
	Sine wave	F3	Full-formant	Sine wave	F3	Full-formant
	Speech Listeners (n=11)					
M	5.90	—	5.66	061	—	2.27
SE	(0.24)	—	(0.33)	(0.10)	—	(0.46)
	Music Listeners (n=11)					
M	10.81	3.97	—	0.04	1.25	—
SE	(7.74)	(0.20)	—	(0.02)	(0.31)	—

Tests for category formation on the sine wave syllables, analogous to those of Experiment 1, showed that stimuli 2 through 5 and 6 through 10 fell into distinct categories for the speech listeners ($p < .05$ for all between-category comparisons; not significant for any within-category comparisons). For the music listeners, there was no effect of stimulus item, indicating that the slope of their mean function did not differ significantly from zero.

Discussion

The results replicate and strengthen those of Experiment 1. The shallower slopes for the speech listeners on sine wave than on full formant syllables indicate, not unexpectedly, that their responses were less consistent for the phonetically impoverished and unfamiliar syllables of the sine wave dialect than for standard synthetic speech. Otherwise, they treated the two forms of speech identically, as the phonetic coherence hypothesis predicts. The music listeners, on the other hand, despite their practice with the F3 tones before hearing the sine wave syllables, performed even less consistently than in Experiment 1. They categorized isolated F3 tones with fair consistency, but they were quite unable to exploit this supposed cue when they heard it in the context of the two lower tones.

Nonetheless, the possibility remained that the group differences might be attributable to the use of labels *per se*. People have had so much more experience with naming words and syllables than with labeling nonspeech sounds, particularly with labeling slight differences in the onset properties of single notes within a chord, that this experiential difference alone might account for the relatively poor performance of the music listeners on the sine wave syllable task. Moreover, the memory demands of the task, which requires listeners to remember the category exemplars in order to label individual items from the continuum, may be much greater for the music listeners than for the speech listeners.

EXPERIMENT 3

A categorization task that does not require labels and that provides the subject with category exemplars on each trial would circumvent the difficulties noted above, as would a discrimination task. We therefore conducted a third study, using an AXB categorization procedure that provides clear-case exemplars on each trial, but does not require category labels (e.g., Bailey, Summerfield, & Dorman, 1977; Best et al.,1981), as well as an AXB discrimination task that places relatively low demands on short-term memory (Best et al., 1981).

Method

Subjects

Thirty-four young adults were tested in the third experiment. The speech bias condition was run on 12 subjects (6 males, 6 females), and the music bias condition on 22 subjects (13 males, 9 females). All had normal hearing and a negative personal and family history of language or speech difficulties, and each was paid $4 for participation. Subsequently, on the basis of posttest questionnaire answers, 1 female speech subject was eliminated because she did not hear the sine wave syllables as speech. Nine subjects were eliminated from the music condition: 8 of these had begun to hear the sine wave syllables as words or syllables (6 males, 2 females), and the remaining subject (a female) had failed to perceive the full-formant series as /la/ and /ra/. The final samples were therefore unequal, with *ns* of 11 and 13 for the speech and music groups, respectively.

Stimuli

The stimuli were the same as in the first two experiments.

Procedure

The subjects were tested under the same listening conditions as before. All subjects completed two tests on each of the three stimulus continua: an AXB categorization test and an AXB discrimination test. They completed the categorization test before the discrimination test for each continuum, in the order (1) sine wave syllables, (2) isolated F3 tones, and (3) full-formant syllables. As in Experiment 1, the sine wave syllable test was presented first to prevent any possible influence of exposure to the other stimulus series on performance with the sine wave syllables. The sine wave syllable test was preceded by the appropriate instruction and training set for the condition randomly assigned to each subject.

On each trial of the AXB categorization tests, three stimuli were presented. The first and third stimuli were constant throughout the test: the endpoint /ra/ or rising F3 item from the appropriate continuum, and the second /la/ or flat F3 item. The middle stimulus, X, varied randomly among the ten items of the stimulus series. The subject's task on each trial was to indicate whether X belonged in the same perceptual category as A (first) or B (third). Each AXB categorization test contained 10 blocked randomizations of the trials for the 10 items in the continuum, with 1.5-sec ISIs, 3.5-sec ITIs, and 5-sec IBIs.

In the AXB discrimination tests, three stimuli were also presented on each trial. However, the first and third stimuli (A and B, respectively) were always three steps apart on the appropriate stimulus continuum and varied from trial to trial, while the middle stimulus (X) always matched either A or B. The subject's task on the discrimination tests was to indicate whether X was identical to A or to B. Each discrimination test contained five randomizations of the 28 possible AXB configurations, blocked in groups of 14 trials, with 1.5-sec ISIs, 3.5-sec ITIs, and 5-sec IBIs.

Results

Categorization

The data were treated as in the previous experiments. Figure 5 displays the group categorization functions, and Table 4 lists the mean category boundaries and slopes, with their standard errors. Category boundaries appear to differ across continua, being lowest on the F3 tones for both groups. The slopes on the sine wave syllables are somewhat steeper than in the previous experiments; but they are still shallowest on the sine wave and steepest on the full-formant syllables for both groups. The music listeners again give a shallower slope than the speech listeners on the sine wave syllables, and the pattern of standard errors for this continuum replicates that of the previous experiments.

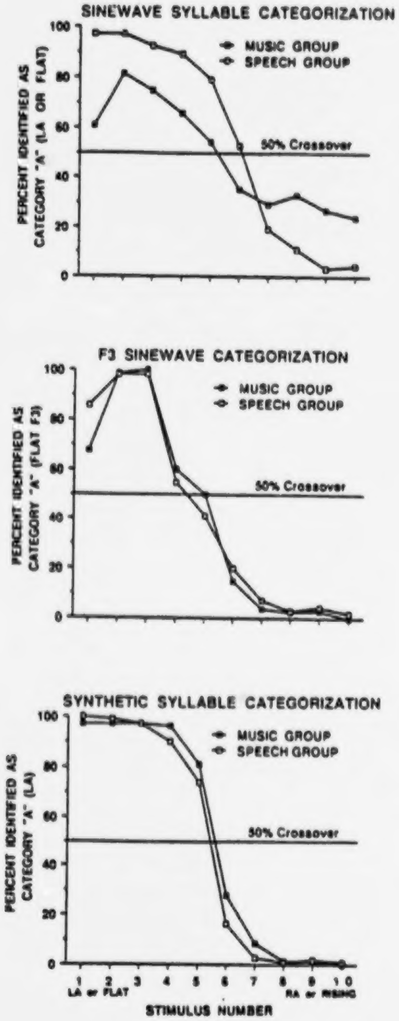


Figure 5. AXB categorization functions for the speech and music listeners on each of the three stimulus continua in Experiment 3.

TABLE 4. Experiment 3: Mean category boundaries and slopes, determined from individual probit analyses on stimulus items 2 through 10, for speech (n11) and music listeners.

	Category Boundaries			Slopes		
	Sine wave	F3	Full-formant	Sine wave	F3	Full-formant
			Speech Listeners (n=11)			
M	6.03	4.82	5.37	0.83	1.24	2.75
SE	(0.19)	(0.27)	(0.17)	(0.10)	(0.32)	(0.90)
			Music Listeners (n=11)			
M	5.19	4.84	5.65	0.26	0.98	1.62
SE	(0.30)	(0.16)	(0.15)	(0.06)	(0.10)	(0.32)

Two-factor (instruction condition × stimulus continuum) analyses of variance, with repeated measures on stimulus continuum, were carried out on category boundaries and slopes. For the category boundaries there was a significant effect of stimulus continuum, $F(2,44) = 7.16$, $p < .002$, a significant interaction between stimulus continuum and instruction condition, $F(2,44) = 3.41$, $p < .04$, but no effect of instruction condition. Scheffé tests on the three pair-by-pair boundary comparisons gave no significant differences for either speech listeners or music listeners.

Analysis of the slope data yielded a significant effect of stimulus condition, $F(2,44) = 9.81$, $p < .0003$, but no effect of instruction condition and no significant interaction. Scheffé tests showed that the sine wave syllable slopes were significantly lower than the slopes for either the F3 tones, $F(1,22) = 10.51$, $p < .05$, or the full-formant syllables, $F(1,22) = 12.56$, $p < .05$, but that the slopes for the full formant syllables and F3 tones did not differ. Although there was no significant interaction, we carried out the planned *t* tests to assess the difference between speech and music listeners' slopes on the sine wave syllables, predicted by the phonetic coherence hypothesis. The difference was again significant, $t(22) = 4.99$, $p < .0001$.

Tests for category formation on the sine wave syllable continuum, analogous to those of the previous experiments, showed a significant effect of stimulus item, $F(8, 80) = 131.43$, $p < .0001$, for the speech listeners, with Stimuli 2 through 5 forming one category and Stimuli 6 through 10 another, according to Scheffé tests ($p < .05$ for all between-category comparisons; not significant for any within-category comparisons). For the Music listeners, there was also a significant effect of stimulus item, $F(8,96) = 16.58$, $p < .0001$; Scheffé tests indicated that Stimuli 1 through 5 fell into one category, Stimuli 7 through 10 into another ($p < .05$ for all between-category comparisons; not significant for any within-category comparisons).

Discrimination

Each subject's percent correct performance was computed for each stimulus pair in each AXB discrimination test. Figure 6 displays mean performance for the two groups on the three continua. The full-formant syllables yield a standard speech continuum pattern for both groups: performance peaks on discrimination pairs that straddle or abut the category boundary. We see a similar, though somewhat flattened, function for the speech listeners on the sine wave syllables. By contrast, the music listeners show no systematic peaks on the sine wave syllables: performance declines across the continuum, as it does for both groups on the F3 tones, although the latter elicit a generally higher level of discrimination than the other continua.

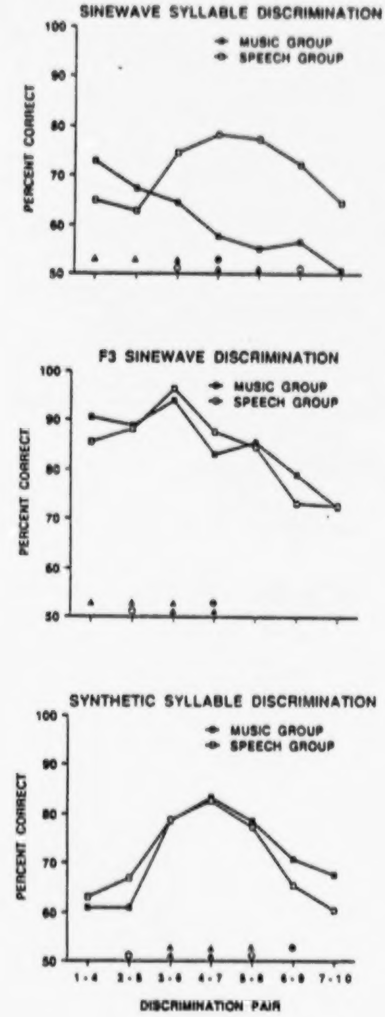


Figure 6. Discrimination functions for the speech and music listeners on each of the three stimulus continua in Experiment 3. The markers above the abcissa indicate the status of each stimulus pair with respect to categorization judgments. Triangles indicate pairs that straddle the category boundary for a given group. Circles indicate pairs in which one item is at or near the boundary. Filled markers represent the music listeners' data, open markers the speech listeners data.

A three factor (instruction condition × stimulus continuum × discrimination pair) analysis of variance, with repeated measures on stimulus continua and discrimination pairs, was carried out. A significant stimulus continuum effect, $F(2,44) = 51.99$, $p < .00001$, indicated that discrimination performance was highest overall for the isolated F3 tones and lowest for the sine wave syllables. A significant interaction between instruction condition and stimulus continuum, $F(2,44) = 5.56$, $p < .01$, combined with Scheffé tests, indicated that the two groups differed in performance only on the sine wave syllable test, on which the speech listeners outperformed the music listeners, $F(1,22) = 13.88$, $p < .05$. Further Scheffé tests showed that performance on the sine wave and full-formant syllable tests did not differ for the speech listeners, but did differ significantly for the music listeners, $F(1,12) = 75.58$, $p. < 001$.

The effect of discrimination pair was significant, $F(6,132) = 16.29$, $p < .00001$, indicating that overall performance level was not uniform throughout the stimulus continua, but rather showed higher performance on some discrimination pairs than on others. A significant interaction between discrimination pair and stimulus continuum, $F(12,264) = 5.29$, $p < .00001$, indicated further that the pattern of the discrimination function differed among the three stimulus continua. A three-way interaction of instruction group with discrimination pair and stimulus continuum, $F(12,264) = 3.01$, $p < .005$, evidently arose because, according to a Scheffé test of the instruction by discrimination pair interaction, $F(6,132) = 6.01$, $p < .01$, only the speech group showed a peak in performance level near the category boundary on the sine wave syllable test.

Discussion

The influence of attentional mode on perception of the sine wave syllables demonstrated in the previous experiments was replicated in Experiment 3. Evidently the difference in the way speech and music listeners categorize these syllables reflects a true difference in perceptual response, and is not simply a function of differences in ability to assign labels to speech versus nonspeech stimuli, nor in the influence of short-term memory for the speech versus nonspeech category exemplars. Although removal of the requirement for overt labeling, as well as presentation of category exemplars for comparison with the target item on each trial, certainly permitted the music listeners to form somewhat more consistent sine wave syllable categories than were observed in the previous experiments, these listeners were still less consistent than the speech listeners, and gave no evidence of a peak at their sine wave syllable category boundary in the discrimination test. Evidently their categories, such as they are, are less robust and more dependent on experimental conditions than those of the speech listeners. Thus, support for the phonetic coherence hypothesis, over the alternative psychoacoustic hypotheses, was considerably strengthened by the third experiment.

GENERAL DISCUSSION

The results of these three experiments are inconsistent with the claim that speech perception entails the simple extraction, or the extraction and integration, of discrete information-bearing elements or cues. All listeners could correctly classify, within psychophysical limits, the transitions on the isolated F3 tones as steady or rising. However, music listeners, biased to listen for the transition in the context of the lower F1 and F2 tones, could not then reliably recover the target pattern. They also could not either integrate the F3 cue with other cues in the F1-F2 array or apprehend the auditory coherence of the total pattern so as to arrive at a unitary, distinctive percept for each category. By contrast, listeners biased to hear the sine wave patterns as speech were evidently immune to whatever psychoacoustic interactions blocked consistent judgments of the patterns by music listeners, in that the former classified the sine wave syllables only somewhat less consistently than they classified the full-formant syllables. These results agree with those of several other studies of sine wave speech in arguing for a specialized mode of speech perception (e.g., Best et al., 1981; Tomiak, Mullenix, & Sawusch, 1987; Williams, 1987).

How are we to characterize this mode? What did the speech listeners in these experiments do that the music listeners did not? Consider, first, the music listeners' performance with the sine wave syllables. In Experiment 1, and particularly in Experiment 3, where labeling was not required so that listeners could compare whole signals without attempting to isolate distinctive cues, the music group's categorization function sloped in the "correct" direction. At least some of these listeners grasped certain contrastive properties of the signals, even though, according to the post-test questionnaire, they did not perceive them as speech. One suspects that, with sufficiently prolonged training under suitable experimental conditions (such as those provided by AXB categorization, as used in Experiment 3), these listeners might even come to render judgments of the sine wave syllables no less consistent than those of the speech listeners. But if they did so, it would remain notable that they require extensive training whereas the speech listeners require very little. Moreover, even if extensive training aided the music listeners, would they then be perceiving the patterns as speech? The answer would surely be yes, if they could tell us the names of the sounds they had heard—that is, if they had discovered the articulatory patterns implied by the signals. But if they could not tell us the names, the answer would be no. Their condition might, in fact, be much like that of non-human animals trained to distinguish between speech sound categories (Kluender, Diehl, & Killeen, 1987). Alternatively, they might be categorizing the patterns on adventitious nonspeech properties, rather as a color-blind individual might correctly classify two objects of different colors on the basis of their differences in brightness rather than of their differences in hue.

Consider here another class of listener well-known to have difficulty with the English /r/-/l/ distinction: monolingual speakers of Japanese. Figure 7 displays a group categorization function for 7 such speakers attempting to classify full-formant patterns on an /r/-/l/ continuum similar to the continuum of the present experiments (MacKain, Best, & Strange, 1981). The function is remarkably like that of the music listeners attempting to classify the corresponding sine wave syllables: some of these Japanese listeners also seem to have captured certain contrastive auditory properties of the signals. However, unlike the music listeners, who were diverted from hearing the sounds as speech by being trained to attend to an acoustic cue rather than to the whole pattern of which it was a part, these Japanese listeners were diverted because they had not discovered the relevant properties of contrastive sounds spoken in an unfamiliar language. However, we know that, with sufficient exposure to the English /r/-/l/ contrast in natural contexts where it serves a phonological function, Japanese listeners can come to hear the contrast correctly and categorically (MacKain et al., 1981). What have they then learned, or discovered? What, more generally, has any second language learner—or, indeed, any child learning a first language—discovered, when the auditory patterns of a target language drop into phonological place? Presumably, they have learned to do more than classify auditory patterns consistently. They have also discovered the correct basis for classification, namely, the articulatory structures that the patterns specify.

Notice that this formulation attempts to resolve the notorious lack of correspondence between a quasi-continuous acoustic signal and its abstract linguistic predicates by positing an event with observable, physical content—the articulatory gesture—as the fundamental unit of both production and perception. We are thus dealing with patterns of movement in space and time, accessible to treatment according to general principles of motor control (Browman & Goldstein, 1986; Kelso, Tuller & Harris, 1983; Saltzman & Kelso, 1987). The task for perceptual theory is then to uncover the acoustic properties that specify a particular pattern of gestures. These acoustic properties will not simply be a collection of independent cues, arrived at by (articulatorily) unconstrained manipulation of a terminal analog synthesizer, but sets of correlated properties that arise from coordinated patterns of gesture, tested by systematic articulatory synthesis.

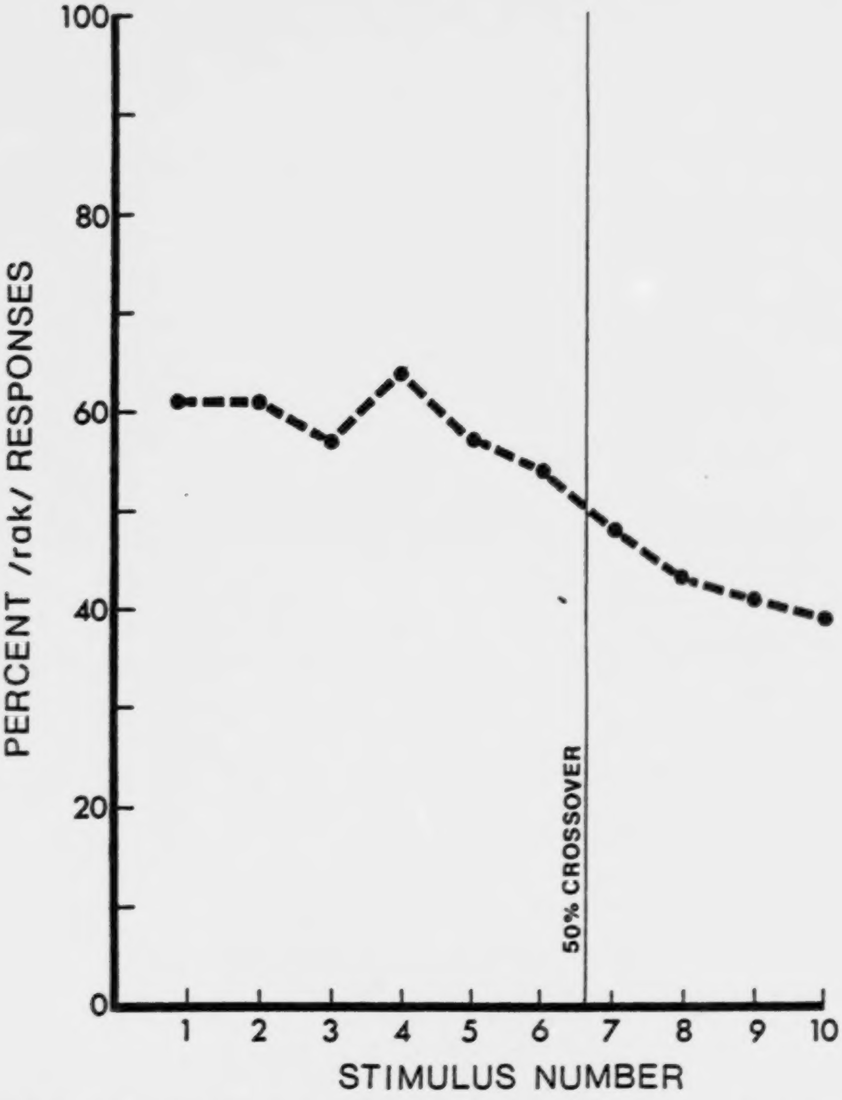


Figure 7. Labeling of a rock-lock continuum by Japanese adults without extensive English conversation experience. Adapted from "Categorical Perception of English /r/ and /l/ by Japanese Bilinguals" by K. S. MacKain, C. T. Best, and W. Strange, 1981, *Applied Psycholinguistics*, 2, p. 378. Copyright 1981 by Cambridge University Press. Reprinted by permission.

We should emphasize that we are not here arguing for a "motor theory" of speech perception. We are not proposing that the *process* of arriving at a speech percept engages mechanisms outside the auditory system that humans share with many other animals. What is specific to speech, and to humans, is the final percept, a phonological structure determined by the structure of the articulatory gestures from which the signal arose (cf. Repp, 1987). Since (as the Japanese listeners show) this structure is inaccessible to adult humans, unless they can assimilate the sounds to the phonological categories of a language they know (see Best, McRoberts, & Sithole, 1988), we assume *a fortiori* that it is also inaccessible to the infant who does not yet know a language. The infant's task is to discover the phonetic coherence of phonological categories in the surrounding language by focusing attention on recurrent auditory contrasts that signal changes of meaning in that language (see Jusczyk, 1986; Studdert-Kennedy, 1986, 1987). An articulatory "representation" of the phonologically contrastive patterns is then an automatic consequence of the species-specific perceptuomotor link that underlies the child's capacity to imitate speech patterns, and so to learn to talk.

Finally, although we have couched our experimental procedures in terms of attention, we do not mean to imply that the processes of speech perception can be engaged or disengaged at will. For, although some listeners can choose to attend or not to attend to particular aspects of a speech signal, such as a speaker's "accent," or even the spectral properties of a fricative noise (Repp, 1981), it is difficult, if not impossible, to hear natural speech, spoken clearly in a language that we know, as nonspeech. And, as we have already remarked, listeners who have once heard a particular sine wave pattern as speech find it difficult later to hear that pattern—or any other sine wave pattern modeled on speech—as entirely devoid of phonetic structure. Evidently, perceiving speech-like patterns as speech is as mandatory, automatic and unconscious as, say, perceiving the rhythm and melody of a nonspeech auditory form.

ACKNOWLEDGMENT

Preparation of this paper was supported in part by NIH Grant HD-01994 to Haskins Laboratories. We thank Len Katz for statistical advice, Michael Dorman, Peter Jusczyk, and Bruno Repp for useful comments on the texts of earlier versions.

REFERENCES

Abramson, A. S., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. In D. E. Commins (Ed.), *Proceedings of the 5th International Congress of Acoustics (A51)*. Liege: G. Thone.

Bailey, P. J., & Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 536-563.

Bailey, P. J., Summerfield, Q., & Dorman, M. F. (1977). On the identification of sine-wave analogues of certain speech sounds. *Haskins Laboratories Status Report, SR51/52*, 1-25.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics, 29*, 191-211.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345-360.

Bregman, A. (1981). Asking the "what for" question in auditory perception. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 99-118). Hillsdale, NJ: Lawrence Erlbaum Associates.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219-252.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review, 81*, 348-374.

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review, 83*, 114-140.

Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Scociety of America, 76*, 1636-1647.

Delgutte, B. (1982). Some correlates of phonetic distinctions at the level of the auditory nerve. In R. Carlson & B. Granstrom (Eds.), *The representation of speech in the peripheral auditory system* (pp. 131-149). New York: Elsevier Biomedical Press.

Dorman, M. F., Cutting, J. E., & Raphael, L. (1975). Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 121-129.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception and Psychophysics, 22*, 109-122.

Finney, D. J. (1971). *Probit analysis*. Cambridge: Cambrideg University Press.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123-136). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jusczyk, P. W. (1986). Toward a model of the development of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 1-19). Hillsdale, NJ: Erlbaum Associates.

Jusczyk, P. W., Smith, L. B., & Murphy, C. (1981). The perceptual classification of speech. *Perception and Psychophysics, 30*, 10-23.

Kelso, J. A. S., Tuller, B., & Harris, K. (1983). A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), *The production of speech*. New York: Springer-Verlag.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science, 237*, 1195-1197.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369-390.

Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.). *The cognitive representation of speech* (pp. 415-420). Amsterdam: North Holland.

Mattingly, I. G., Liberman, A. M., Syrdal, A. M., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology, 2*, 131-157.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*, 172-191.

Pastore, R. E. (1981). Possible psychoacoustic factors in speech perception. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1980). Speech perception without traditional speech cues. *Science, 212*, 947-950.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception and Psychophysics, 30*, 217-227.

Repp, B. H. (1987). The role of psychophysics in understanding speech perception. In M. E. H. Schouten (Ed.), *The Psychophysics of Speech Perception* (pp. 3-27). Boston: Martinus Nijhoff.

Saltzman, E., & Kelso, J. A. S. (1987). Skilled actions: A task-dynamic approach. *Psychological Review, 94*, 84-106.

Schouten, M. E. H. (1980). The case against a speech mode of perception. *Acta Otolaryngologica, 44*, 71-98.

Studdert-Kennedy, M. (1986). Sources of variability in early speech development. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes*, Hillsdale (pp. 58-76). NJ: Lawrence Erlbaum Associates.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 67-84). London: Academic Press.

Tomiak, G. R., Mullenix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Scociety of America, 81*, 755-764.

Williams, D. R. (1987). The role of dynamic information in the perception of coarticulated vowels. Unpublished doctoral dissertation, University of Connecticut.

FOOTNOTES

[*] *Perception and Psychophysics*, 1989, 45(3), 237-250.

[†] Also Wesleyan University.

[††] Also University of Connecticut and Yale University.

[†††] Massachusetts Institute of Technology.

[††††] NYNEX Science and Technology, White Plains, NY.

[1] In this paper we use the term *cohere*, and its derivatives, to refer to the effect of a perceptual process by which listeners apprehend a complex signal as a unitary pattern, or configuration, rather than as a collection of discrete elements. In vision, we may add to the many examples familiar from textbook treatments of Gestalt principles, the phenomenon of face recognition, where the identity of a face emerges as a holistic pattern, not simply a collection of discrete features. In speech, the unitary patterns would correspond to units of linguistic function, such as phonemes, syllables, morphemes, words.

[2] Because the first stimulus on the continuum had a slightly falling transition (originally intended to help bias the full-formant series toward /l/, as noted above), listeners (particularly the music listeners on the sine wave syllables, and both groups on the F3 tones) tended to judge this stimulus with slightly less consistency than its neighbors (see Figures 2 through 5). As a result, probit analyses tended to yield lower slopes than were characteristic of the main bodies of the functions, and so to exaggerate the slope differences between groups and conditions. We therefore omitted this stimulus from the probit analyses: all computed means and slopes in this and the following experiments are based on analyses of individual functions for stimuli 2 through 10.

Note, further, that by converting the standard deviations of the underlying distributions into their reciprocals (the slopes of the cumulative functions), we went some way toward homogenizing the group variances, as appropriate for subsequent analyses of variance. At the same time, we reduced the apparent differences between groups across stimulus continua. For example, in Table 1 the difference between the mean slopes for the two instruction conditions is not much greater on the sine wave syllables (0.80 - 0.14 = 0.66) than on the full formant syllables (2.30 - 1.70 = 0.60). But the difference between the mean standard deviations for the two conditions is very much greater on the sine wave syllables (1/0.14 - 1/0.80 = 5.89) than on the full formant syllables (1/1.70 - 1/2.30 = 0.15). The reader should bear this in mind when comparing slopes depicted in the figures with slope values listed in the tables.

[3] One of the five music subjects from Experiment 1a, who returned to take the speech test, later turned out to be the one we rejected from that experiment because she had heard some of the sine wave syllables as "sounding like r." However, her data were not appreciably different from those of other music listeners, and so we retained her in Experiment 1b. If her tendency to hear some of the sine wave syllables as r-like had facilitated her categorization of these syllables, the result would presumably have been to reduce the differences between instruction conditions. However, the results of statistical analyses on the data of Experiment 1b were unchanged when this fifth subject's data were eliminated.

Reception of Language in Broca's Aphasia*

Donald Shankweiler,† Stephen Crain,† Paul Gorrell,†† and Betty Tuller†††

This experiment tests between two competing hypotheses about the source of failures in comprehension by Broca-type aphasics with agrammatic production. These are characterized as 1) the hypothesis that these aphasic individuals have sustained a partial loss in syntactic knowledge, and 2) the hypothesis that, despite intact structural knowledge, they suffer from inability to put that knowledge to use in comprehension tasks such as object manipulation and sentence-picture matching. To decide between the hypotheses, this study compared the speed and accuracy of Broca-type aphasics with a control group of normal subjects using an on-line grammaticality judgement task in which the anomaly involved closed-class vocabulary items. The results are in accord with the view that the source of agrammatic performance is not a loss of syntactic knowledge, since the responses of the aphasic group closely mirror those of the control group (e.g., word position effects were found for both groups). The results are interpreted, instead, as support for the alternative view that agrammatic aphasics have difficulties in processing syntactic knowledge.

INTRODUCTION

Our concern in this research is with a pattern of symptoms that is characteristic of Broca's syndrome, the most notable feature of which is a marked reduction in fluency. The utterances of the persons we studied consist for the most part of short phrases, often haltingly and effortfully produced and with considerable constriction of vocabulary. It is untenable to suppose that these symptoms are the result purely of motor interference with the execution of fluent speech because the omission of words is selective. The utterances of our subjects typically consist of nouns and verbs with relatively few grammatical words such as particles, auxiliary verbs and prepositions. This telegraphic manner of speaking has often been called "agrammatic" (Isserlin, 1922; Pick, 1913) to underscore the selective nature of the omissions: the words that are omitted in the affected individual's productions are more often the grammatical (or "closed class") words of the sentence (articles, prepositions, auxiliary verbs, etc.) than the content (or "open class") words. Omission of these items results in sentences that are not well-formed. Thus the concept of agrammatism carries with it the idea that at the core of Broca's syndrome is a disturbance in the structure of spoken language. Alternatively, these individuals might retain the critical structures but suffer from inability to put the knowledge to use due to a limitation in processing. As a step toward developing a satisfactory understanding of the aphasias, we carried out an experiment to test between these competing hypotheses.

Originally, the concept of agrammatism was used to designate a disorder of speech output, but as a result of more recent developments in research on aphasia it is arguable that the deficit that underlies this condition is in fact more general, such that both production and comprehension are often impaired (Berndt & Caramazza, 1980; Bradley, Garrett & Zurif, 1980; Caramazza & Zurif, 1976; Goodglass, Gleason, & Hyde, 1970; Heilman & Scholes, 1976; Parisi & Pizzamiglio, 1970; but see also Miceli, Mazzucchi, Menn, & Goodglass, 1983, for an indication that comprehension deficits are not always associated with agrammatic production). The comprehension impairments that are often part of the symptom picture in aphasia of the nonfluent type may escape notice because the aphasic individual usually appears to understand what the examiner says. Practical judgment and reasoning capacities are often remarkably intact, allowing the deficit in grammatical processing to remain hidden.

In a pioneering attempt to isolate the source of the language comprehension deficit associated with Broca's aphasia, Zurif, Caramazza and Meyerson (1972) applied a metalinguistic judgment procedure in which aphasic subjects were asked to indicate which words in a sentence belong together. These investigators concluded that Broca-type subjects failed to use syntactic categories in making their judgments, using other grouping strategies instead. In later research, Caramazza and Zurif (1976) tested comprehension of syntactically complex sentences containing relative clauses by means of a picture matching task. It was found that patients diagnosed as Broca's aphasics performed poorly on implausible sentences and sentences that were semantically reversible, such as (1).

(1) The lion that the tiger is chasing is fat.

In the absence of context, the meaning of such sentences cannot be inferred by appeal to world knowledge. Together, the findings led the authors to propose that their Broca subjects were "asyntactic" in comprehension. They attributed the apparently intact comprehension displayed by these individuals in ordinary social communication to unimpaired "heuristic" strategies: the aphasic subject simply combines the meanings of the individual lexical items in a plausible way, thereby compensating for the loss of syntactic ("algorithmic") analysis. Caramazza and Zurif took a large step toward identifying the source of the comprehension disturbances in Broca's aphasia, but it remained an open question how these difficulties were related to the impairments in production.

A linguistic account which explicitly ties together the production and comprehension impairments was proposed by Bradley and her colleagues (Bradley, 1978; Bradley, Garrett, & Zurif, 1980). This proposal focused on the role that the closed-class morphology plays in models of normal language processing.[1] According to the Bradley et al. model, the closed class portion of the lexicon can be accessed by two routes, but only one of these is ordinarily used in parsing sentences. In agrammatism, that route is selectively impaired, although agrammatic aphasics still retain closed class items in their mental lexicon. This dual route hypothesis permits Bradley et al. to suppose that agrammatics have lost only the ability to use the closed class vocabulary in constructing phrasal structure for sentences. But as a consequence of disruption to the mechanism that accesses this vocabulary, there is impairment of syntactic operations both in production and perception. The idea of double representation of closed-class words in the lexicon has been challenged on empirical grounds (see Gordon, 1983; Gordon & Caramazza, 1982). Though the notion of bifurcation of the lexicon may have to be abandoned, Bradley and her colleagues attempted to provide a compact explanation for the deficits that commonly co-occur in Broca-type aphasia: both production and comprehension failures were seen as manifestations of impairment of some of the procedures for lexical retrieval.

It is worth noting that the proposal of Bradley et al is consistent with the possibility that syntactic knowledge is retained in Broca-type aphasics who have comprehension difficulties. Difficulties would arise if the normal routes of access to stored linguistic knowledge were blocked. Alternatively, the source of the difficulty could be at some post-syntactic level. Data addressing this issue were obtained by Linebarger, Schwartz and Saffran (1983a). Dissatisfied with the ambiguity in interpreting the results of commonly-used tests of comprehension, these investigators asked whether Broca's aphasics are capable of constructing syntactic representations by having them judge the grammaticality of auditorily presented sentences. They found evidence of preserved sensitivity to transformational operations, subcategorization of verbs, and at least partial sensitivity to the syntactic functions of some closed-class vocabulary items. Since the aphasic subjects studied by Linebarger et al. had been found to fail with regularity on sentence-picture matching and acting out tests, their success with grammaticality judgments came as a surprise to adherents of the asyntactic hypothesis. The findings seemed to fly in the face of previous research and theorizing which had sought to establish that a phonological or syntactic deficit was responsible for the sentence processing difficulties of agrammatic aphasics. If the judgment task accurately reflects syntactic competence, then some component of language processing at a higher level than the syntax must be responsible for errors that these individuals make on standard measures of sentence comprehension.[2]

In view of the importance of the questions raised by the findings of Linebarger et al., the theory and methodology of comprehension testing in aphasia is deservedly undergoing thorough scrutiny (Caplan, 1985; Caramazza & Berndt, 1985; Kolk, Van Grunsven, & Keyser, 1985). Predictably, the conclusion that Broca's aphasics have retained syntactic competence has not gone unchallenged (Caplan & Futter, 1986; Grodzinsky, 1986; Zurif & Grodzinsky, 1983). The debate hinges on the problem of inferring competence from various kinds of test data. In deciding how claims about the source of comprehension failures should be evaluated, we must ask which tasks give an accurate indication of where the problems lie. Although criteria for an adequate test are rarely specified, we suggest two. First, the task must allow one to draw inferences about a single level of processing. Comprehension tasks must be capable of teasing apart the individual contributing factors in order to pinpoint the source of the problem within the language apparatus. A second criterion requires the task to reflect the routine operations of the language understanding system. Let us see how existing comprehension tests measure up to these requirements, which we will call the *univocality* criterion and the criterion of *naturalness*.

These two criteria are often in competition, because we are trying to disentangle factors that are intertwined in ordinary spoken communication. For this reason, some commonly used methods for assessing comprehension fail to meet one or both criteria. For example, the method of sentence-picture matching seems to satisfy the criterion of naturalness, but it fails the criterion of univocality because it confounds syntactic and semantic processing. The method of object manipulation confounds a third factor with these two. In addition to carrying out a structural analysis of a test sentence, a subject must formulate a plan for demonstrating his understanding (see Hamburger & Crain, 1984, 1987, for data and discussion of the role of the planning stage in sentence comprehension). The univocality criterion will not be met in an object manipulation task whenever the plan associated with a sentence is as complex as either its syntactic or semantic structure, since failure to perform accurately in this circumstance cannot be unequivocally interpreted. This task also frequently fails the naturalness criterion, because subjects are asked to act out sentence meanings in the absence of prior context. This renders the task inappropriate for any construction that bears presuppositional content, since, by their nature, presuppositions must be satisfied before an utterance is made (for a discussion of how the failure of researchers to satisfy presuppositions has masked children's knowledge of syntax, see Hamburger & Crain, 1982).

Let us now apply these criteria to the grammaticality judgment task used by Linebarger et al. (1983a). Consider first the criterion of univocality: the semantic and planning stages are largely circumvented in this task since the subject has only to indicate whether or not the test sentences are grammatical, but there may be sources of ambiguity inherent in this technique, too. Unless proper controls are built into the test sentences, there is no way of knowing which linguistic level of representation the subject is appealing to in making the judgment. When a sentence is made ungrammatical, even by the substitution of a single word, the sentence that results may be semantically anomalous as well as syntactically deviant. Also, the substitution may alter the prosody, creating an anomaly which the subject could conceivably detect without performing a structural analysis at all. However, if sentence meaning and prosody are adequately controlled, the grammatical judgment task would presumably meet the criterion of univocality.

The second criterion for evaluating comprehension tasks is naturalness: does the grammaticality judgment task reflect normal sentence processing? In a reply to Linebarger et al. (1983a), Zurif and Grodzinsky (1983) challenge the grammaticality judgment task in assessing the syntactic knowledge of aphasics on the grounds of naturalness. They contend that "given the luxury of conscious reflection permitted by acceptability judgments, one might expect not to tap normal on-line processes." They go on to speculate that the judgment task employed by Linebarger et al. (1983a) would allow an agrammatic subject ... "to use processing routes other than those by which structural information is normally made available at the appropriate time for interpretation (p. 208)." Thus, in their view, success on this task could draw upon metalinguistic abilities that stand apart from the usual operations of the language understanding mechanism.

Resolution of the issue of naturalness hinges on what characteristics are imputed to the human sentence parsing mechanism. Various conceptions of how the parser works have been proposed, but many questions remain open. For example, no consensus exists concerning whether the parser operates in a serial or in a parallel manner (Frazier & Fodor, 1978; Gorrell, 1987), or whether its mode of operation is deterministic or nondeterministic (Fodor, 1985; Marcus, 1980;). There is widespread agreement on one characteristic, however. As Zurif and Grodzinsky suggest, it is widely accepted that the operations of the parser are carried out "on line," allowing the rapid construction of syntactic and semantic structure as the input string is received.

While we know of no explicit definition of on-line processing, we consider three laboratory phenomena to be indicators that the parser constructs a representation as the sentence unfolds. First, there is evidence from a variety of sources, including the present study, that decisions about both syntactic and semantic structure are made before the perceiver has heard all the words of a sentence (e.g., Crain & Steedman, 1985; Marslen-Wilson, 1975). Second, there is evidence that processing load increases as the distance increases between words that form discontinuous constituents, such as verb/particle constructions or agreement in number between a determiner and noun. Discontinuity is seen to tax working memory resources by postponing the closure of syntactic categories. Setting aside a constituent to await other elements with which it is to be associated is generally assumed to be costly of memory resources (e.g., Wanner & Maratsos, 1978). This is consistent with the claim that the parser works most efficiently when it can form constituents locally, on a word by word basis.

Evidence of on-line processing is also provided by many detection experiments which find that the identification of a target segment—a specific phoneme, letter, or word—is faster as a sentence progresses. We call this the *word position* effect.[3] The word position effect probably reflects the normal on-line operation of the parsing mechanism. In the course of parsing, the perceptual processing system assigns structure to a sentence without waiting until all its words or constituents have arrived, allowing lower level information to be rapidly segmented and shunted upward through the system. This enables the processor to work on several levels in parallel within a narrow "window" of only a few words (e.g., Frazier & Fodor, 1978; Marcus, 1980). Later items are detected faster because assignment of a structural analysis reduces the number of subsequent structural options that otherwise would be pursued. This, in turn, frees up computational resources needed for other processing tasks. Since information in verbal working memory can be stored only briefly before it decays (Baddeley & Hitch, 1974; Conrad, 1964), the availability of a mechanism for the rapid on-line integration of incoming information is essential for the interpretation of extended speech or text.

It is surprising, in view of the interest the question has generated, that the available data on comprehension in agrammatism leaves open the question of how well agrammatics can process items in the closed-class vocabulary in tasks that tap normal (on-line) sentence parsing operations. The experiment of Linebarger et al. (1983a) was not designed to assess the on-line processing capabilities of their subjects. Nonetheless, any attempt to dismiss the grammaticality judgment task on *a priori* grounds is unwarranted, since even *post hoc* judgments could in principle reflect structural analyses computed by the subject on line. The purpose of the present investigation was to study the question of whether agrammatic and normal speakers use implicit knowledge of closed-class words in the same way, both in computing syntactic structure for a given input string and in making on-line decisions about grammaticality. In designing the stimulus materials and procedure, we have been mindful of the criticisms raised by Zurif and Grodzinsky because much rides on the correct interpretation of the Linebarger et al. (1983a) findings.

The present experiment satisfies the criterion of naturalness in so far as it was designed to reveal whether agrammatics assign syntactic structure on line. We developed a chronometric technique to test the speed and accuracy with which normal control subjects and aphasics can detect grammatical errors in sentences in which grammaticality turns on the use of a closed-class vocabulary item. The technique converts the ordinary grammaticality judgment task into one that taps on-line processing: It achieves this, first, by presenting each sentence only once at a normal conversational rate; second, by emphasizing speed of decision, requiring a response as soon as the violation is detected, thus allowing little opportunity for reflection; and, third, by controlling the location of the word that normal subjects would identify as the source of ungrammaticality of each test sentence.

The chief goal of this investigation was to discover whether the receptive language impairment commonly associated with "agrammatic" production results from impaired grammatical knowledge involving the closed class morphology or whether the impairment is caused by a processing difficulty. This question is framed against a backdrop of syntactic theory—as well as a current proposal concerning the nature of linguistic processing (e.g., the Modularity Hypothesis of Forster, 1979, and Fodor, 1983).[4] Within the framework of Government-Binding theory (Chomsky, 1981), Grodzinsky (1984a) has proposed that the tendency of some aphasics to omit closed-class items in production, as well as their failures to interpret them correctly, is due to a selective impairment at the level of S-structure. On Grodzinsky's hypothesis, both the production and comprehension problems of agrammatic aphasics reflect impaired knowledge at this level of representation: what is lost in agrammatism is the lexical content normally present at the terminal nodes of closed-class vocabulary items in the construction of a representation at the level of S-structure.[5] Items belonging to *different* syntactic categories should be preserved, despite loss of sensitivity to distinctions *within* a category. So, for example, if an adverb were inserted in a sentence at a terminal node that requires a determiner, the agrammatic subject would detect this substitution, and would find the sentence anomalous. We will call this a between-class substitution. An example is given in (2).

(2) *The cabdriver forgot to bring the senator to *away* rally.

It follows from Grodzinsky's hypothesis, however, that a substitution *within* the same form class would not be noticed, even if this substitution leads to grammatical violations. A within-class substitution of this kind occurs in sentence (3). For sentences like this, agrammatic aphasics would simply have to guess whether the correct item was present under the auxiliary node, as in (4).

(3) *Peter *have* planning to see a new movie Saturday night.

(4) Peter *was* planning to see a new movie Saturday night.

To test Grodzinsky's proposal, the ungrammatical sentences in the present study contained violations of lexical insertion both within and between classes. Since agrammatics cannot "look" under the lexical node in (3) any more than in (2), they might respond in the same way to both sentence types. On the other hand, a subject who is overaccepting would respond correctly (although for the wrong reasons) to sentences like (3), but not to sentences like (2), which contain a grammatical violation.

Method

Subjects

In the present study, our strategy is to compare performance of normals with nonfluent aphasic subjects who have been selected using liberal criteria. We required only that the aphasics satisfy an operational definition of agrammatism by evincing significant limitations in production of closed-class vocabulary items.

Six normal subjects and six aphasic subjects participated in the experiment. All were native speakers of English with normal hearing. All were fully ambulatory and able to travel to the testing site. The characteristics of the aphasic subjects are summarized in Table 1. In each case, the aphasic symptoms were attributable to stroke which had occurred from 3 to 14 years prior to testing. Hospital records dating from the acute phase were scrutinized. The diagnostic workup included CAT scan for all patients except LS. The clinical neurological examination revealed right hemiplegia in each case and the scan data indicated infarction in the territory of the middle cerebral artery, predominantly frontal, but with varying degrees of posterior extension into the Sylvian region. All the aphasic subjects had received diagnostic aphasia testing prior to the present study, including portions of the Boston Diagnostic Aphasia Examination. All were judged able to understand the task directions.

As may be seen from Table 1, each aphasic subject showed frequent omission of closed-class vocabulary items and marked reduction in of speech output. Their responses to the "cookie theft" picture are agrammatic also in the sense that most of the sentences are not well formed. With regard to comprehension, however, the subjects show less homogeneity. Only four of the six subjects showed significant impairment on comprehension of passive voice sentences. Two of the group comprehended both sets nearly without error, and one of these was perhaps the most severely affected of all the subjects in production.

TABLE 1

APHASIC SUBJECTS: BACKGROUND DATA

SUBJ	AGE	EDUC	ETIOLOGY	COMPREHENSION BDAE (Z Score)	COMPREHENSION ACTIVE (Percent Correct)†	COMPREHENSION PASSIVE (Percent Correct)†	PRODUCTION BDAE COOKIE THEFT PICTURE
*PJ	47	12	L.Frontal Embolism (1979)	1.0	98	98	Mom is washing her... dishes and her ... dishes. But I can't say it. Dishes going... this way. I can't say it. Her... dishes is uh... fall and goes down there.
*ME	56	16	Infarct, L. Mid Cerebral Artery (1979)	0.8	98	98	Cookies dar, no I... cookie jar. Tilts, tilts, stool. Wash dishes, I, no, she is... washing dishes. She pour... she... she pour water. (Where?) She pour...pour water floor....
*VS	61	12	Infarct, L. Mid Cerebral Artery (1970)	0.7	88	60	Mother is washing the...dishes... The David and Kathy ... cookies. Fall on. The... the... water. The water is... clump... flowing. The chair is... the boy is counter.
*LS	52	12	Occlusion, L. Mid Cerebral Artery (1974)	0.4	67	52	A mother...a dish...drying... Plate...a faucet...running... a boy, eating cookies... eating the cookies... girl.
AK	70	4	Infarct, L. Mid Cerebral Artery (1979)	0.7	79	42	Mother... washing dishes. The boy... falling. Eat..s..s... coo...kies. Girl cookie. Mother..s..s... not look.
ED	68	12	Infarct, L. Mid Cerebral Artery (1980)	0.2	71	29	House... kitchen... woman clean dishes. The sink is full... water... spilling all... over.

*ME and PJ also participated in Schwartz et al. (1987).
*VS and LS also participated in the study of Linebarger et al. (1983).
†These percentages are based on A', a nonparametric index of sensitivity. The materials consist of semantically reversible active and passive sentences, (24 of each). Each was tested twice, once with a matching picture and once with a picture in which agent-patient roles are reversed. A' is theoretically equal to the proportion of correct responses attainable in a two-alternative forced-choice procedure.

The normal subjects were adult speakers of English who were matched approximately for age (mean 51 years) and education (mean 14 years) with the aphasic group. We will refer to them as the reference group.

Materials

The stimulus set consists of 112 sentences of ten words each, half of which are ungrammatical. Forty of the ungrammatical sentences contain a single grammatical violation involving a closed-class item. These, together with their matched grammatical control sentences, are listed in the Appendix.[6] The grammatical violations were of four kinds; each involved a determiner, preposition, particle or verbal element. The four categories were equally represented. In each anomalous sentence a word (or a bound morpheme) occurred which, in that context, made the sentence ungrammatical. We call this the *target*. For example, one type of violation involved agreement in number between a determiner and a noun, as in sentence (5). Here, the target "word" is the plural morpheme *s*.

(5) The banker noticed that two customer<_> deposited the checks late.
Control: The sailor noticed that three shipmates left the dock early.

Detection of the ungrammaticality in (5) requires access to the grammatical feature (singular/plural) associated with the determiner and the affix of the noun. We will refer to the determiner (DET) in this case as the *licensing* word since the recognition of its features is essential to detection of the anomaly, even though it is not itself the word that we expected people to alter to make the sentence grammatical (i.e., it is not the target word). In contrast, in violations involving particles (PART) such as (6), the licensing word is an open class item, the verb "picking," whereas the target is a closed-class item, the particle "on."

(6) Picking the birthday present *on* would be nice for Susan.
Control: Picking an expensive gift out will be pleasant for George.

For half of the sentences the substituted item preserves category membership (e.g., an inappropriate preposition for an appropriate one); for half it does not (e.g., a preposition for a determiner). For every ungrammatical sentence there is included a grammatical (control) sentence matched with it in length and structure, as illustrated above. Also controlled is serial position of the target word, i.e., whether it occurs near the beginning (within the group consisting of words 1 - 3), middle (words 4 - 7), or near the end of the sentence (words 8 - 10). An additional factor is the proximity of the closed-class target item to the word which determines its appropriateness (e.g., a determiner and a head noun). If the licensing word and the target word are adjacent, the proximity is CLOSE, as in (5); if separated by at least two intervening words, it is FAR, as in (6).

The test sentences were tape-recorded at normal conversational speed and the waveforms were digitized and stored. The test tape was prepared with the aid of a waveform editing program at Haskins Laboratories (WENDY). The procedure was as follows: for each ungrammatical sentence, a pitch pulse was placed in the nonspeech channel to coincide with the onset of the word that caused the ungrammaticality (the *target* word). In order to accomplish this, the waveform for each target word was displayed with expansion of the time scale such that the onset of voicing could be located reliably. A timing pulse was placed at that point in each of the ungrammatical sentences. This procedure allows us to discover if the subject is responding before hearing *all* the words in the sentence. The timing pulse triggered a response-time clock (accurate to 1 msec) which was stopped by a subject's key press. (For the grammatical control sentences, a pulse was placed at the onset of voicing of the initial word in the sentence. Since correct responses to grammatical sentences must always occur after the last word, no further use was made of response-time data for these sentences.)

Care was taken to exclude extraneous cues that might offer a spurious basis for judgment by permitting the subject to capitalize on a prosodic or semantic anomaly. Such extra-syntactic factors were controlled for as follows: First, the possibility of inadvertently creating ungrammatical test sentences with anomalous prosody, which might cue the correct response without syntactic processing, was minimized by recording and presenting each target only once, at a normal conversational rate, and by avoiding unnatural discontinuities that occur when tape splicing is used. Secondly, a trained speaker was employed in preparation of the stimulus tape. The speaker was instructed to read each ungrammatical sentence as though it contained a grammatically appropriate word in place of the anomalous word, thereby creating a sentence token that, though ungrammatical, conformed to normal prosodic contours. In order to avoid confounding syntactic analysis with sensitivity to semantic and pragmatic factors, the sentences were designed so that the meaning of the sentence was transparent and semantically plausible despite the presence of an anomalous word.

Another source of possible variability needs to be addressed. Each of the ungrammatical test sentences could be rendered grammatical by changing a single word, namely the word we designate as the target word. But, of course, each sentence could also be corrected in numerous other ways. For example, even the simple sentence *She were writing a letter* can be corrected as *She was writing a letter*, or *They were writing a letter*, and so on. Therefore, one cannot assume without argument that subjects will always identify a particular word as the anomalous item. But just this assumption is critical for our analysis of on-line processing, for example, regarding the interpretation of the word-position effect.

A review of the relevant research indicates that people are remarkably consistent in the word they identify as anomalous in ungrammatical sentences like those we constructed for the present study (Crain & Fodor, 1987; Freedman & Forster, 1985). Returning to the example given above, this research leads us to expect that subjects would consistently correct the sentence to yield: *She was writing a letter*. Correcting the sentence in this way suggests that speakers of English typically pursue a strategy which, proceeding from the left, accepts as many words as can be incorporated into a single structural representation. A perceiver who is pursuing this strategy locates the anomaly on the last alternative possibility. This led us to expect a consensus among our subjects in how the anomalous test sentences would be construed.

We validated this expectation empirically by asking 20 undergraduate students to correct the anomaly in each test sentence, to discover whether or not they would correct each sentence by changing the target word and that word alone. The results of the correction task revealed that 93% of the responses were either alterations or deletions of the target word. By extension, we can suppose that the aphasic subjects and the reference group of the present experiment display a comparable degree of consistency and that their judgments reflect a common diagnosis of what is wrong with each of the anomalous sentences.

Procedure

Subjects were tested individually in a sound-treated room. They indicated their responses to each test sentence by a key press. The response keys were labeled with cartoon faces: a smiling face to indicate a grammatical sentence, and a frowning face to indicate an ungrammatical one. The subject's task was simply to press the appropriate key as rapidly as possible on hearing each test sentence. Subjects were instructed to press the frowning face key as soon as they detected an ungrammaticality, and not to wait until the end of the sentence. A practice session consisting of 12 sentences, with examples of each type, was used to ensure that each subject understood the task.

Scheme of the data analysis

Statistical treatment of the measures of response time and accuracy for the ungrammatical (target) sentences was carried out to assess the between-subjects factor of *group* (aphasic vs. reference group) and the four within-subjects (structural) factors that were controlled in the test sentences. These are listed below:

(i) *position*: the ordinal position of the critical closed-class word in the sentence, i.e., BEGINNING vs. MIDDLE vs. END.[7]

(ii) *substitution type:* whether the substituted closed-class item preserves category membership, i.e., WITHIN- vs. BETWEEN-class violations.

(iii) *proximity:* the proximity of the closed-class item to the licensing word, i.e., CLOSE vs. FAR.

(iv) *category*: four categories of closed-class words are represented determiners (DET), prepositions (PREP), particles (PART), and agreement of verbal items with their antecedents (AUX).

Results

Initially, the data from both the normal subjects and the aphasic subjects were combined, and parallel ANOVAs were carried out on the response-time and accuracy measures for the ungrammatical sentences. As one might anticipate, each structural variable had a significant effect on reaction time. Likewise, all but one had a significant effect on error rate (only *position* failed to reach significance). Considering first the response time measures, the results of the analysis for the combined data from both groups are as follows: *position*, $F(2,22) = 21.79$; $p < .01$; *proximity*, $F(1,11) = 22.50$; $p < .01$; *substitution type*, $F(1,11) = 4.83$; $p < .05$, and *category*, $F(3,33) = 10.40$; $p < .01$. With regard to errors, three of the four variables were again significant. The results are as follows: *position*, $F(2, 22) = 2.04$; $p < .15$, *proximity*, $F(1,11) = 9.18$; $p < .01$, *substitution type*, $F(1,11) = 8.27$; $p < .02$, and *category*, $F(3,33) = 2.81$; $p < .05$. We consider the interpretation of these results presently when we discuss further analyses that test for differences between the subject groups.

Table 2 summarizes the accuracy of responses by subject. The top half of the table gives the percent correct for the ungrammatical sentences and their grammatical controls for the aphasic subjects: the corresponding data for the reference group are given in the bottom half. It may be seen that five of the six aphasic subjects performed with a higher level of accuracy on the grammatical control sentences (column 5) than on the sentences containing grammatical violations (column 2). The means were 80% and 75% correct, respectively, so the level of performance was roughly equivalent for both sets of sentences. In no subject was the discrepancy greater than 11%. In the absence of any marked tendency for the aphasic subjects to be overaccepting, further analyses are confined to the errors on the ungrammatical sentences. For the reference group, the mean performance level was 97% correct on both the ungrammatical sentences and grammatical control sentences.

TABLE 2

PERCENT CORRECT BY SENTENCE TYPE FOR INDIVIDUAL SUBJECTS

APHASIC GROUP

SUBJECT	OVERALL	WITHIN-CLASS	BETWEEN-CLASS	CONTROL
PJ	83	80	85	88
ME	90	79	100	96
VS	55	35	75	56
LS	75	60	90	83
AK	78	55	100	73
ED	68	60	75	79
MEAN	75	61	88	80

REFERENCE GROUP

SUBJECT	OVERALL	WITHIN-CLASS	BETWEEN-CLASS	CONTROL
BC	90	85	95	92
KB	100	100	100	100
CS	97	95	100	92
EW	100	100	100	100
TM	100	100	100	100
JM	97	100	95	96
MEAN	97	97	98	97

In addition, Table 2 partitions scores on ungrammatical sentences in which the target word preserves or violates category membership— WITHIN-class (column 3) vs BETWEEN-class (column 4). It should be noted that all six of the aphasic subjects and five of the six members of the reference group were more accurate in detecting between-class substitutions. The import of this result will be considered presently.

The major task of the data analysis was to test between the possibility that the aphasics' lower accuracy and longer response latencies reflect loss of syntactic structures or, alternatively, whether they point to a deficit in linguistic processing. This question is examined by testing for interaction between the experimental variables and group. We reasoned that if the aphasic subjects have incurred a breakdown in the internal grammar that blocks full syntactic analysis of closed-class words, we could expect to find consistent between-group interactions on either or both of the dependent measures, response time and errors (where an error is failure to detect an ungrammatical sentence).

By contrast, the hypothesis that Broca's aphasics suffer from limitations in processing, but have preserved sensitivity to syntactic structure, would lead us to anticipate a different pattern of results, i.e., main effects for within-subject variables but no between-group interactions except those that arise as artifacts of measurement. Spurious interaction would result, for example, whenever the reference group reached a ceiling level of performance by responding with nearly perfect accuracy in making grammaticality judgments. Because the findings involving response time and accuracy parallel each other closely, we discuss both aspects of the results together, rather than allocating them to separate sections.

Figure 1 displays the performance of each group with respect to each of the four structural variables. We will refer to the panels in this figure in the order from upper-left to lower-right. The effects of ordinal *position* of the target word on response time and accuracy are shown in the pair of graphs at the top of the figure.[8] It is apparent that with respect to response time, performance increases with ordinal position for both the aphasic group $F(2,5) = 34.57$; $p < .01$ and the reference group $F(2,5) = 7.47$; $p < .04$. As may be seen from inspection of graph at top left, both groups show slower responses to an anomalous target word at the beginning of a sentence than at the middle or end. The interaction of *group* and *position* is significant $F(2,20) = 10.88$; $p < .01$, and so deserves further comment. The interaction occurred because the slope of the position effect is considerably steeper for the aphasic subjects than for the reference group, but the effect of position for both groups is in the same direction. With regard to accuracy, there was no significant effect of position for either group, as noted above.

The beneficial effect of increased *proximity* of the closed class item to its licensing word was a significant factor in task performance for both groups, as shown in the second pair of graphs. It is apparent that there is a decrement in performance for the value FAR, both in speed and accuracy. This decrement is significant for response time, for the aphasic group $F(1,5) = 10.67$; $p < .02$ as well as for the reference group $F(1,5) = 27.05$; $p < .01$. It is noteworthy that there was no interaction with group ($p < .3$). As for accuracy, the effect of proximity was also significant for the aphasic group $F(1,5) = 20.04$; $p < .01$. They responded correctly on 90% of the cases where the licensing word was CLOSE but only 63% when it was FAR, yielding a difference of 27%. For the reference group, the corresponding difference is only 2%, clearly not significant. Because the reference group was performing at ceiling in both conditions, an interaction arises $F(1,10) = 15.41$; $p < ,01$.

A comparison of the individual aphasic subjects' performances on the experimental materials reveals sufficient consistency to justify treating the aphasics as a group and reporting summary statistics on the grouped data, as presented in the foregoing paragraphs. Figure 2 shows the individual subject response times for *word position* and *proximity*, the two factors on which our argument for on-line processing rests. It is noteworthy that the effect of both variables is consistent across the individual subjects, despite fairly marked differences in overall level of performance. Each subject shows monotonically decreasing response times as a function of word position, and each shows faster response times for sentences in which the licensing word was close to the target word.

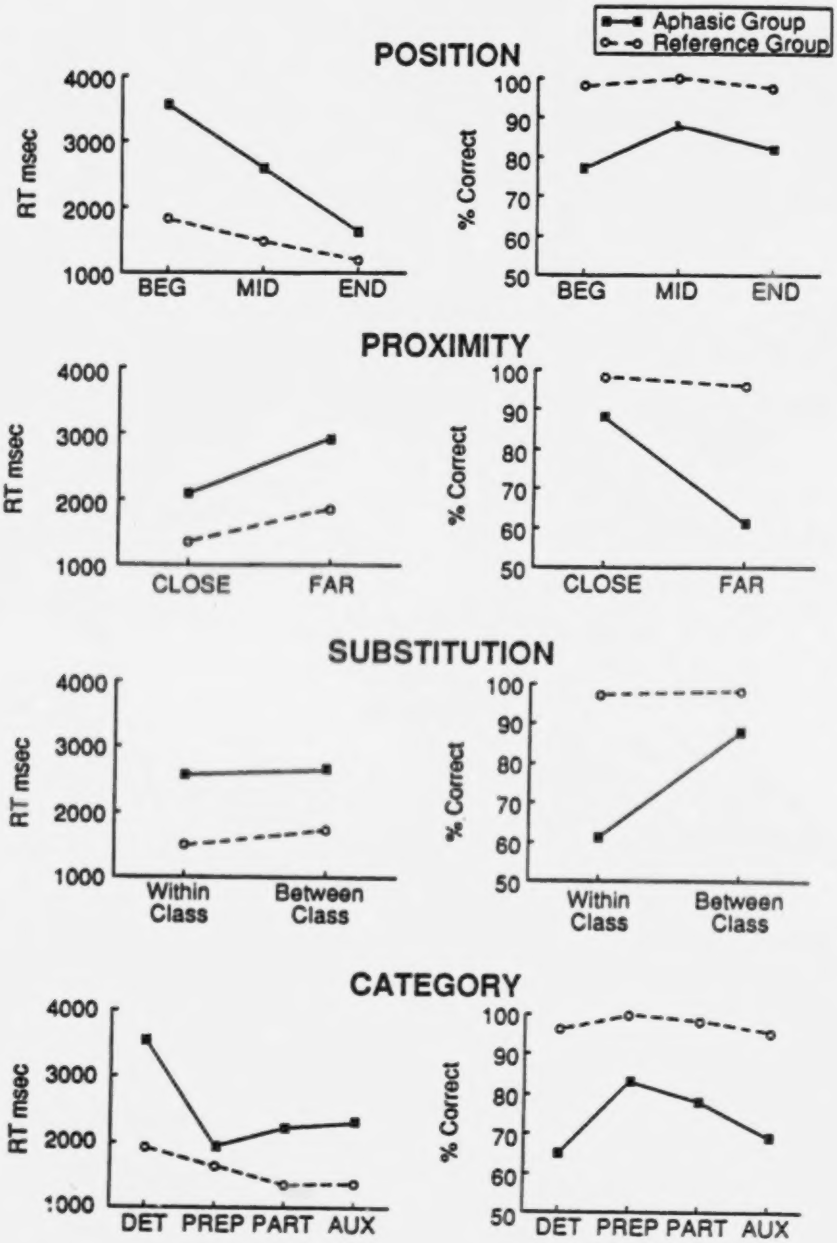


Figure 1. Mean performance of the aphasic group and the reference group on reaction time and percent correct detection of ungrammatical sentences for each structural factor.

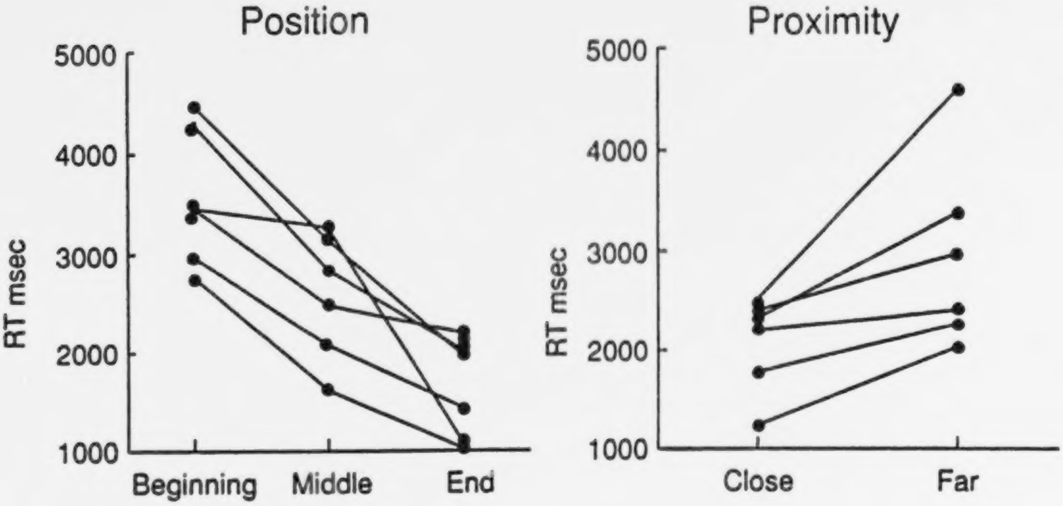


Figure 2. Mean performance of individual aphasic subjects for Position and Proximity.

We have already noted differences in accuracy of responding by *substitution type*. As shown in Table 2, all of the aphasic subjects were more accurate in rejecting the between-class substitutions. The difference was significant for the aphasic group $F(1,5) = 27.25$; $p < .01$, but not the reference group, presumably because the latter were performing at ceiling level on both types of substitutions. The reference group detected 98% and 97%, respectively, whereas the aphasics detected 61% of the within-class and 88% of the between-class substitutions. This led to a significant interaction by group $F(1,10) = 13.68$; $p < .01$.

There was also an overall effect of *substitution type* on response time, but here the pattern was reversed. As shown in the third pair of graphs in Figure 1, the between-class substitutions evoked longer response times for both groups. The effect of *substitution type* on response time proved to be a significant effect for the reference group $F(1,5) = 12.13$; $p < .01$, but not for the aphasic group $F(1,5) = 0.49$; $p < .5$, and there was no interaction of this factor with group. (See Discussion section for comments on the disparity in the findings for response time and accuracy).

The effects on performance of syntactic *category* are shown in pair of graphs at the bottom of Figure 1. Recall that, within the set of test sentences, there were grammatically inappropriate substitutions among four categories of closed-class words: prepositions, particles, verbal elements and determiners. There is a significant effect across these categories on response time (but not on accuracy) for both the aphasic group $F(3,5) = 8.66$; $p < .01$ and the reference group $F(3,5) = 13.81$; $p < .01$. The interaction of *group* and *category* is significant also $F(3,30) = 5.63$; $p < .01$. Further analysis shows that the interaction is due chiefly to one category, prepositions. The relatively better performance of the aphasic group on prepositions is consistent with findings of others (e.g., Friederici, 1982; Grodzinsky, 1984b). With this category omitted from the analysis, the interaction disappears ($p < .2$).

(e.g., Friederici, 1982; Grodzinsky, 1984b). With this category omitted from the analysis, the interaction disappears ($p < .2$).

In summary, the results reveal both similarities and differences in the pattern of responses for aphasic subjects and the reference group. As for the similarities, we take the existence of the word position effect and the proximity effect in both groups to be a telling indication of on-line language processing. Differences between the groups are indicated by the consistently lower accuracy and longer response times of the aphasic subjects and by the occurrence of some interactions of the structural variables with group. In interpreting the interactions it was important to distinguish genuine interactions from spurious statistical effects that result from a problem in measurement arising, for example, when normal subjects approach a ceiling level of performance. In the following section we consider the implications of the findings for understanding the comprehension difficulties that may occur in cases of nonfluent aphasia and also the wider implications for the concept of agrammatism. We begin by examining the specific findings concerning sensitivity of the aphasic subjects to the closed-class vocabulary, using the reference group as a basis for comparison. Then we discuss how the findings fit within the larger framework of current theories of syntactic processing in aphasia.

Discussion

Applying the logic sketched in the Introduction, we looked for evidence of preserved on-line sentence processing by Broca-type aphasics, and we sought to delimit the locus within the language system of their comprehension difficulties. We assessed performance on the test sentences with respect to the four structural variables: *position, proximity, substitution type*, and *category*. It is evident that the findings favor the hypothesis of preserved sensitivity to the syntactic functions of the closed-class vocabulary. This is indicated by critical similarities in response patterns for both accuracy and response latency for the aphasic group and the reference group on each factor. Thus the idea that the comprehension difficulties of the aphasic subjects are due to loss of structures is not supported. Instead, a deficit in processing is implicated. We now discuss the results in detail.

Position. Consider first the effect of the position of the anomalous word in the sentence. Both groups displayed the word position effect, showing a decrease in response time for later-occurring (in contrast to earlier-occurring) anomalous words. In previous work on sentence processing in normals, this effect has been interpreted as a demonstration that language perceivers begin to construct a structural representation of a sentence almost from the outset. If this reasoning is correct, the appearance of the word position effect in the aphasic group constitutes evidence of preserved syntactic competence. The reference group was both faster and more accurate than the aphasic group, as expected, but both groups displayed the same trend.

The finding that the aphasics, like the normals, were consistently faster in detecting violations in late-arriving items calls into question a proposal by Caplan (1982) that agrammatics assign minimal structure to sentences above the lexical level. If we are correct in supposing that the word position effect is attributable to on-line construction of the syntactic structure, then these aphasic subjects must compute enough structure above the lexical level to enable the input string to be processed as a structured entity. It is apparent, then, that some Broca's aphasics are capable of using information about phrase structure above the lexical nodes as the basis for grammaticality decisions, as Linebarger et al. (1983a) claimed.

Proximity. Further evidence that the grammaticality judgment task involves on-line sentence processing is the beneficial effect of proximity of the licensing and the target words.[9] If a sentence is processed on-line, the difficulty of relating two lexical items should increase with the distance between the items. It is presumably more difficult to relate two items if other items intervene than if they are adjacent because the grammatical features of the first item must be retained in working memory. On the other hand, if the sentence were being processed off line its entire representation would be held in memory and no local effects of differences in proximity would be expected.

Proximity of the closed-class item to its licensing word was a significant factor in task performance for both subject groups. Each group performed significantly better for the grammatical anomalies in which the distance of the closed-class item to its licensing word was CLOSE. When there was greater distance between the licensing word and the target item, the aphasics, like the normal subjects, were slower to respond. It is noteworthy that the effect of proximity, like the effect of word position, was in the same direction for all the aphasic subjects. The finding that both groups had more difficulty in the FAR condition is an indication that the more remote dependencies stress working memory. This, together with the effect obtained for ordinal position of the anomalous word, lends support to the conclusion that both groups processed the test materials on line.

Substitution-type. A noteworthy feature of the design of this study permitted a comparison of the accuracy of detection of within-class and between-class substitutions. This comparison is of special interest in view of the recent theoretical claims of Grodzinsky (1984b; 1986) to which we have referred. Let us review the findings. First, we noted a main effect on response time for substitution type, with the between-class substitutions evoking longer response times, but no interaction of this factor with *group*. Both groups took longer to respond to between-class substitutions, although the effect did not reach significance for the aphasic group. Substitution type also had a significant effect on accuracy of responding, but the direction of the effect was reversed; the within-class substitutions, as expected, evoked a greater number of errors than the between-class substitutions. This difference was significant for the aphasic group, but not for the reference group, presumably because they were performing at ceiling level on both types of substitutions.

The disparity between the findings for response time and accuracy deserves comment, since it might appear paradoxical that within-class substitutions should evoke more errors, yet were more rapidly identified as incorrect. In our view both results are consistent with the other manifestations of on-line processing. To show this we must first explain why the between-class substitutions took longer; then we must explain why the within-class substitutions, though faster to detect, induced more errors.

In regard to the response time differences on between-class substitutions, we suggest that the reason an ungrammatical variant of an expected word (e.g., *were* instead of *was*) may be easier for the sentence parser to recover from than a totally inappropriate word is as follows: when an item cannot be fit into the current parse of a sentence, the listener is jolted into a new attempt a structuring the available material. Attempts at restructuring are costly of time because there exists no correct alternative syntactic structure, but once the search has been exhausted, listeners would be accurate in rejecting the sentence.

In considering the response times for within-class substitutions, one must bear in mind that the analysis is based only on the data for sentences on which subjects gave the correct, negative response. Therefore, response times to the within-class substitutions reflect only those trials on which the subject was able to see that an illicit variant of an appropriate item had been substituted for it. Since the substitute was obviously inapt in these cases, no attempt at restructuring was made. On the other hand, the high rate of 'misses' for within-class substitutions among the aphasic subjects may reflect an automatic tendency to replace an illicit item with an appropriate one. Thus, the errors that occurred for within-class substitutions may simply represent a confusion between the item that was actually perceived and the spontaneously corrected item that subjects unconsciously replaced it with. Automatic correction of ungrammatical strings was a prominent phenomenon in a profoundly aphasic patient studied by Whitaker (1976). The operation of automatic correction may help to account for the relatively poor performance of aphasics in response to erroneous within-class substitutions. An account in these terms would not implicate the syntactic processor as the source of the problem. Instead, the problem would be seen as one of monitoring and suppressing information that the parser has automatically computed.

Category. The findings concerning accuracy of responses to variations in category also merit comment. Although the aphasic group was significantly less accurate on within-class than between-class substitutions, they performed above chance on two kinds of within-class substitutions: prepositions and particles. The response pattern is relevant to the proposal by Grodzinsky (1984a) and Zurif and Grodzinsky (1983), who maintain that in agrammatism the terminal nodes immediately dominated by the lexical nodes DET, AUX, INFL, COMP, etc. are left unspecified.[10] As Zurif and Grodzinsky (1983) point out, "violations of syntactic structure are permissible in agrammatism (that is, not noticed by the patient) only to the extent that they are either errors involving inflections, auxiliaries and determiners, or omissions of prepositions (p. 210)." Thus, agrammatic subjects should be unable to detect *any* within-class substitution but should remain sensitive to substitutions which fail to preserve category membership.[11] The fact that the aphasic group responded with above-chance accuracy in response to particles (79% correct) and prepositions (83% correct) cannot easily be accommodated by the Zurif and Grodzinsky proposal.[12]

Also recalcitrant for their proposal is the finding that the aphasic group responded accurately to control sentences 80% of the time. As noted earlier, the Zurif and Grodzinsky proposal applies to the control sentences as well as to the ungrammatical target sentences—a subject who could not identify items beneath lexical nodes would err in the same, but opposite, proportion to each sentence type. For example, a subject who judged 80% of the control sentences to be correct should also have judged 80% of the ungrammatical sentences to be correct, using the same criteria. This would have resulted in only 20% accuracy for the ungrammatical items. The findings are inconsistent with this expectation, however. The aphasic subjects were 61% correct in responding to the target sentences. Recall, also, that the target sentences and their controls were identical in structure, except for the critical (anomalous) word, so that any differences in response rates must be attributed to the ungrammaticality of the test sentences.

Having considered the effects of each of the structural variables individually, we can now ask whether the findings form a coherent picture of the processing capabilities of the aphasic subjects. Earlier we discussed three indicators of on-line computation of structural representations: the word position effect, the proximity effect, and the detection of an anomaly before the sentence is complete. If the aphasic subjects are using closed-class vocabulary items in constructing a syntactic representation on line, we would expect them to exhibit each of these phenomena. We have presented evidence that demonstrates a word position effect and an effect of proximity. We found that aside from differences in overall level of performance, both the reaction-time and the accuracy data for each group are concordant (assuming that we are correct in attributing the interactions with subject group to ceiling performance by the normal subjects.) With regard to the remaining indicator of on-line processing, we note that 50% of the responses of the aphasic group were made before the arrival of the last word in the sentence in cases where the grammatical violation was near the beginning. Moreover, the post-sentential response times lagged an average of only 182 ms after the end of the sentence, surely insufficient time for conscious reflection.[13]

Other data relevant to the issue of on-line processing by aphasics are presented in a recent case study by Tyler (1985). Like our aphasic subjects, the nonfluent aphasic studied by Tyler showed a word position effect in monitoring normal prose. However, this subject did not display an effect of word position for semantically anomalous prose. In this respect, Tyler's aphasic subject manifested an aberrant profile—in normal subjects a word position effect shows up in processing both standard text and semantically incoherent, but grammatically correct, sentences. Since the decrement in response time with advancing position is putatively due to the construction of a syntactic structure as the sentence proceeds, Tyler interprets the absence of a decrement in the anomalous prose condition as indicating the subject's inability to construct normal syntactic representations of sentences. This leads her to speculate that the word-position effect displayed by her subject in the standard prose condition might have resulted from an overreliance on semantic/pragmatic information. It should be noted, though, that whatever ability underlies this subject's processing in the standard prose condition, it elicited the usual word position effect, rendering his performance indistinguishable from normal. Tyler offers the interpretation that her aphasic subject suffers from damage to two processors, resulting in deficient syntactic processing coupled with heightened sensitivity to semantics.

Another interpretation of the performance pattern displayed by this aphasic subject can be given—an interpretation that obviates the need to postulate opposing aberrant systems that conspire to make processing appear normal in ordinary circumstances. We raise the possibility that this subject's syntactic and semantic knowledge is intact. That is the simple conclusion to draw from the normal profile in the standard prose condition. But because the subject also has impaired processing capabilities, his ability to analyze language structure on line is weakened. As a consequence, the appearance of semantically incoherent word combinations may impose sufficient stress to disrupt syntactic processing. Put another way, our suggestion is that the syntactic processing capabilities of Tyler's subject are adequate under ordinary conditions, where both syntactic and semantic cues converge on an appropriate analysis. But in adverse conditions syntactic processing may become derailed.

In our aphasic subjects, too, we would maintain that basic syntactic knowledge is intact. Moreover, the findings of the present research show clear indications that sentence processing routines that eventuate in grammaticality judgments are applied on-line. This effectively answers the concerns expressed by Zurif and Grodzinsky (1983). Our results suggest, by extension, that the conclusion reached by Linebarger et al. (1983a) of preserved sensitivity to syntactic structure in Broca-type aphasics is not vitiated by the off-line character of their judgments task. The finding of a word position effect in the present study, which also used judgments of grammaticality, invites the inference that in the Linebarger study, too, these judgments reflect on-line decisions by the subjects.[14] Clearly, our aphasic subjects did not wait to apply syntactic processing routines until after all the words of a sentence had been encountered. If a subject were applying these routines in this way, the appearance of an effect of serial position would be mysterious. In short, there is evidence that Broca's aphasics are capable of on-line sentence processing, and that this processing is reflected in their grammaticality judgments.

Given our concern in this research with the issue of on-line processing in aphasia, the discussion inevitably sought to draw out the parallels in response pattern on the grammaticality judgments test between the aphasic subjects and the normal subjects. The value of this was to rule out a whole class of explanations of the nature of the basic deficit in agrammatic aphasia. What can we say, then, about the deficits which so definitely are displayed by the aphasic subjects? All of the nonfluent aphasics we have studied displayed major limitations in speech production that involve access to the closed-class vocabulary, confirming earlier indications of the critical importance of this portion of the lexicon for agrammatism. And, in four of the six subjects, these difficulties in speech production were accompanied by difficulties in comprehension, as evidenced by their performance on sentence-picture matching to passive-voice sentences. But, as we noted, there are inconsistencies in comprehension; two subjects performed almost without error on the test of passives. We cannot at present explain the apparent disparity between the results of studies using comprehension tasks, which present paradoxical findings concerning sparing and loss of syntactic knowledge, and studies using judgment tasks, which consistently demonstrate sparing. We might add that none of the current theories of agrammatism, the *Mapping Hypothesis* of Linebarger et al. (1983a; Schwartz et al. 1987), or Grodzinsky's *Trace Theory* (1986), can account for such discrepant individual patterns of response on language reception tasks in the face of consistent severe impairment in language production. The inconsistencies mean that a unitary account of agrammatism that cuts across both production and perception of language remains elusive. Future research would do well to focus on subjects who show these discrepancies, and to explore further the comprehension tasks which have led to inconsistent response patterns.

In summary, we conclude that the locus of the deficits in our aphasic subjects is in the processing system and not the system that represents grammatical knowledge. Were we to attempt to maintain that the findings reported for the aphasic group are the result of loss of syntactic representations, we would be left without an explanation for the basically similar performance of the reference group. The proposal that the locus of the aphasics' deficit is in the processing system allows us to explain the parallels they showed with normal subjects as a reflection of spared syntactic competence. At the same time, the differences between the groups—longer response times and lower accuracy rates for the aphasic group—can be seen as reflections of the reduced processing capability that the aphasic subjects bring to the task of constructing a syntactic structure for an input string. We would predict, then, that if normal subjects were to perform a similar task under adverse listening conditions (e.g., time compressed speech (Chodorow 1976) or the rapid serial visual presentation technique of Forster (1970)), differences between the groups in response times and accuracy rates would narrow. We are currently investigating this possibility.

APPENDIX

Experimental Sentences Containing Grammatical Violations and Matching Grammatical Control Sentences

AUX

Within-class substitutions

1) Roger wants the tusks that two large African elephants *gives*.
Beth wants the wool that one big mountain sheep gives.

2) The baker told the helper that the bread *were* rising.
The farmer told the stranger that a tornado was coming.

3) Peter *have* planning to see a new movie Saturday night.
Mark is training to win the big race next week.

4) The Indian said that several bass *was* getting very large.
The hunter said that one deer is getting awfully scared.

5) One of the sheep *were* sipping water from the trough.
Two of the fish were eating worms off our hooks.

Between-class substitutions

1) One of the faculty *the* selling tickets for the dance.
Four of the bison were chasing hunters across the field.

2) The mechanic told the driver that the fender *there* dented.
The waiter told the diner that the pies were fresh.

3) The clerk said that six fish *here* getting quite hungry.
The farmer said that two sheep were getting closely sheared.

4) Harry *that* attempting to reach the tall branch all day.
George was going to visit his best friend on Sunday.

5) Paul wants the fillet that *on* tasty Alaskan salmon give.
Tom wants the battle that many huge rainbow trout give.

DET

Within-class substitutions

1) Several *man* begged to be allowed into the movie theater.
Several men begged to be admitted to the murder trial.

2) The realtor remembered to visit a newly listed country *houses*.
The teacher remembered to bring two large clean glass jars.

3) The banker noticed that two *customer* deposited the checks late.
The sailor noticed that three shipmates left the dock early.

4) Bill thinks many of the best *tool* disappeared last night.
Terry thinks one of the worst players left this morning.

5) The astronaut forgot to bring the compass to a *spaceships*.
The bellboy forgot to bring the luggage to the room.

Between-class substitutions

1) The cabdriver forgot to bring the senator to *away* rally.
The paratrooper forgot to bring the map to the plane.

2) Sam thinks *out* of the old factories exploded after midnight.
Harry thinks several of the new lambs arrived before lunch.

3) The carpenter noticed that *on* homeowner bought the wood yesterday.
The lawyer noticed that one client paid the bill late.

4) Above *child* begged to be admitted to the football stadium.
A youngster begged to be invited to the birthday party.

5) The old seamstress remembered to donate *what* ragged cotton dresses.
The young salesman remembered to take a heavy wool jacket.

PREP

Within-class substitutions

1) The machinist pointed to the lathe that Max fell *out*.
*The plumber pointed to the wrench that Allen fell over.

2) Out *up* the street an old truck was approaching noisily.
Out of the window an engineer was waving his arm happily.

3) The janitor pointed to the broom that Phil swept *of*.
*The gardener pointed to the sprinkler that Larry tripped over.

4) The milkman was speaking *out* a man who needed advice.
A policeman was talking to a woman who needed directions.

5) Of *the* overgrown snowy woods a rancher carried his saddle.
Across the parched sandy desert a soldier was carrying his canteen.

Between-class substitutions

1) The mailman was looking *must* a dog that needed food.
The fireman was shouting at a child who needed help.

2) Away *here* a goal a soccer coach was running wildly.
Away from the plane a stunt pilot was walking jauntily.

3) The electrician pointed to the wire that Chunk fell *what*.
*The plumber pointed to the wrench that Allen fell over.

4) Never *the* grassy rolling hills a family carried their picnic.
Through the dense evergreen forest a lumberjack carried his axe.

5) The operator pointed to the switchboard that Barbara worked *where*.
*The gardener pointed to the sprinkler that Larry tripped over.

PART

Within-class substitutions

1) The wife of the owner took the new hobby *down*.
A friend of the manager took a big sign out.

2) The chef reheated the stew that the customer pushed *beside*.
The programmer designed the game that the company brought out.

3) Picking the birthday present *on* would be nice for Susan.
Picking an expensive gift out will be pleasant for George.

4) The good natured baker put *at* a white floppy hat.
The well mannered butler put out a large grey cat.

5) Making *on* funny songs is difficult for choir directors.
Making up long grocery lists is boring for busy housewives.

Between-Class substitutions

1) Making *then* ghost stories is hard for young children.
Making out bad checks is easy for clever criminals.

2) Picking a tasteful tie *the* could be simple for ministers.
Picking a suitable prize can be fun for winners.

3) The pitcher delivered the ball that the batter swung *now*.
The ranger relit the fire that the campers put out.

4) The child of the driver took a tiny truck *was*.
The mother of the thief took the shiny jewel away.

5) The well attired banker put *is* a fine silk scarf.
The well groomed barber put on a new white coat.

Note: Sentences preceded by * serve as controls for more than one ungrammatical target sentence.

ACKNOWLEDGMENT

This research was supported in part by a Program Project Grant to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994). We thank Myrna Schwartz for allowing us to study L.S., V.S., and M.E., and Eleanor Saffran for allowing us to study P.J. We are grateful to both, and to Marcia Linebarger, for their encouragement of this project and for kindly providing background data on their subjects. We are indebted to Weijia Ni for help in preparation of the manuscript.

REFERENCES

Aaronson, D. (1968). Temporal course of perception in an immediate recall task. *Journal of Experimental Psychology, 76*, 129-140.

Badecker, K., & Caramazza, A. (1985). On considerations of method and theory governing the use of clinical categories in neurolinguistics and cognitive neuropsychology: The case against agrammatism. *Cognition, 20*, 97-125.

Baddeley, A. D., & Hitch, G. B. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press.

Berndt, R., & Caramazza, A. (1980). A redefinition of Broca's aphasia: Implications for a neuropsychological model of language. *Applied Psycholinguistics, 1*, 225-278.

Bradley, D. (1978). *Computational distinctions of vocabulary type*. Unpublished doctoral dissertation, MIT.

Bradley, D., Garrett, M., & Zurif, E. (1980). Syntactic deficits in Broca's aphasia. In D. Caplan (Ed.), *Biological studies of mental processes*. Cambridge, MA: MIT Press.

Caplan, D. (1982). Syntactic competence in agrammatism: A lexical hypothesis. In M. Studdert-Kennedy (Ed.), *The neurobiology of language*. Cambridge, MA: MIT Press.

Caplan, D. (1985). Syntactic and semantic structures in agrammatism. In M.-L. Kean (Ed.), *Agrammatism*. New York: Academic Press.

Caplan, D. (1986). In defense of agrammatism. *Cognition, 24*, 263-276.

Caplan, D., & Futter, C. (1986). Assignment of thematic roles to nouns in sentence comprehension by an agrammatic patient. *Brain and Language, 27*, 117-134.

Caramazza, A., & Berndt, R. (1985). A multicomponent view of agrammatic Broca's aphasia. In M.-L. Kean (Ed.), *Agrammatism*. New York: Academic Press.

Caramazza, A., & Zurif, E. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language, 3*, 572-582.

Chodorow, M. (1976). *Experimental studies of syntactic and lexical processes in language comprehension*. Doctoral dissertation, MIT.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris Publication.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 3*, 75-84.

Crain, S., & Fodor, J.D. (1987). Sentence matching and overgeneration. *Cognition, 26*, 123-169.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing*. Cambridge: Cambridge University Press.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. D. (1985). Deterministic parsing and subjacency. *Language and Cognitive Processes, 1*, 3-42.

Forster, K. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception and Psychophysics, 8*, 215-221.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W. Cooper & E. Walker (Eds.), *Sentence processing: Studies presented to Merrill Garrett*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Foss, D. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior, 8*, 457-462.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Doctoral dissertation, University of Connecticut.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291-325.

Freedman, S., & Forster, K. (1985). The psychological status of overgenerated sentences. *Cognition, 19*, 101-131.

Friederici, A. (1982). Syntactic and semantic processes in aphasic deficits: The availability of prepositions. *Brain and Language, 15*, 249-258.

Goodglass, H., Gleason, J. B. & Hyde, M. R. (1970). Some dimensions of auditory language comprehension in aphasia. *Journal of Speech and Hearing Research, 13*, 595-606.

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior, 22*, 146-160.

Gordon, B., & Caramazza, A. (1982). Lexical decision for open- and closed-class words: Failure to replicate differential sensitivity. *Brain and Language, 15*, 143-160.

Gorrell, P. (1987). *Studies of human syntactic processing: Ranked-parallel versus serial models*. Doctoral dissertation, University of Connecticut.

Grodzinsky, Y. (1984a). *Language deficits and linguistic theory*. Doctoral dissertation, Brandeis University.

Grodzinsky, Y. (1984b). The syntactic characterization of agrammatism. *Cognition, 16*, 99-120.

Grodzinsky, Y. (1986). Language deficits and the theory of syntax. *Brain and Language, 27*, 135-159.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj (Ed.), *Language development Volume 1: Syntax and semantics*. Hillsdale, NJ: Lawrence Erlbaum Associates,

Hamburger, H., & Crain, S. (1984). Acquisition of cognitive compiling. *Cognition, 17*, 85-136.

Hamburger, H., & Crain, S. (1987). Plans and semantics in human processing of language. *Cognitive Science, 11*, 101-136.

Heilman, K., & Scholes, R. (1976). The nature of comprehension in Broca's, conduction and Wernicke's aphasics. *Cortex, 12*, 258-265.

Holmes, V., & Forster, K. (1970). Detection of extraneous signals during sentence recognition. *Perception and Psychophysics, 7*, 297-301.

Isserlin, M. (1922). Ueber agrammatismus. *Zeitschrift fuer die Gesammte Neurologie und Psychiatrie, 75*, 332-416.

Kean, M.-L. (1977). The linguistic interpretation of aphasic syndromes: Agrammatism in Broca's aphasia, an example. *Cognition, 5*, 9-46.

Kean, M.-L. (1980). Grammatical representations and descriptions of language processors. In D. Caplan (Ed.), *Biological studies of mental processes*. Cambridge, MA: MIT Press.

Kean, M.-L. (1982). Three perspectives for the analysis of aphasic syndromes. In M. Arbib, D. Caplan, & J. Marshall (Eds.), *Neural models of language processes*. New York: New York Press.

Kolk, H., van Grunsven, M., & Keyser, A. (1985). On parallelism between production and comprehension in agrammatism. In M.-L. Kean (Ed.), *Agrammatism*. New York: Academic Press.

Lapointe, S. (1983). Some issues in the linguistic description of agrammatism. *Cognition, 14*, 1-39.

Linebarger, M., Schwartz. M., & Saffran, E. M. (1983a). Sensitivity to grammatical structure in so-called agrammatic aphasia. *Cognition, 13*, 361-392.

Linebarger, M., Schwartz. M., & Saffran, E.M. (1983b). Syntactic processing in agrammatism: A reply to Zurif and Grodzinsky. *Cognition, 15*, 215-225.

Lukatela, K., Crain, S., & Shankweiler, D. (1988). Sensitivity to inflectional morphology in agrammatism: Investigation of a highly inflected language. *Brain and Language, 33*, 1-15.

Marcus, M. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science, 189*, 226-228.

Marslen-Wilson, W. (1984). Function and process in spoken word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language process*. Erlbaum: Hillsdale, NJ.

Marslen-Wilson, W., & Tyler, L. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1-72.

Miceli, G., Mazzucchi, A., Menn, L., & Goodglass, H. (1983). Contrasting cases of Italian agrammatic aphasia without comprehension disorder. *Brain and Language, 19*, 65-97.

Parisi, D. & Pizzamiglio, L. (1970). Syntactic comprehension in aphasia. *Cortex, 6*, 202-215.

Pick, A. (1913). *Die agrammatische Spachstoerungen*. Berlin: Springer.

Schwartz, M., Linebarger, M., & Saffran, E. M. (1985). The status of the syntactic deficit: The theory of agrammatism. In M.-L. Kean (Ed.), *Agrammatism*. New York: Academic Press.

Schwartz, M., Linebarger, M., Saffran, E. M., & Pate, D. S. (1987). Syntactic transparency and sentence interpretation in aphasia. *Language and Cognitive Processes, 2*, 85-113.

Sproat, R. (1986). Competence, performance, and agrammatism: A reply to Grodzinsky. *Brain and Language, 27*, 160-167.

Tyler, L. (1985). Real-time comprehension processes in agrammatism: A case study. *Brain and Language, 26*, 259-275.

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Whitaker, H. (1976). A case study of the isolation of the language function. In H. Whitaker & H. A. Whitaker (Eds.), *Studies in neurolinguistics* (Vol. 2). New York: Academic Press.

Zurif, E., & Grodzinsky, Y. (1983). Grammatical sensitivity in agrammatism: A reply to Linebarger et al. *Cognition, 15*, 207-213.

Zurif, E., Caramazza, A., & Meyerson, R. (1972). Grammatical judgments of agrammatic aphasics. *Neuropsychologia, 10*, 405-417.

FOOTNOTES

[*] *Language and Cognitive Processes*, in press.

[†] Also University of Connecticut.

[††] University of Maryland.

[†††] Center for Complex Systems , Florida Atlantic University.

[1] An alternative proposal by Kean (1977) argues that the core problem in agrammatism is in the phonological component. Kean supposed that components of the language apparatus are hierarchically organized, with a unidirectional flow of information between levels. Thus, a disturbance at a lower level in the system (i.e., the phonology) could masquerade as a breakdown at a higher level. In this way Kean attempted to explain the nature of the apparent "agrammatic" output associated with nonfluent aphasia. She maintained that the Broca's aphasic tends to reduce the structure of a sentence to the minimal string of elements which can be lexically construed as phonological words. This idea can be extended to account for both their production and comprehension impairments: The Broca's aphasic fails to produce or attend to suffixes and other elements which do not occur independently as phonological words—elements which are critical for subsequent syntactic and semantic analysis. Kean's is an elegant attempt to apply linguistic theory to explain a broad set of phenomena in aphasia. However, this viewpoint runs into the difficulty that Broca's aphasics are apparently sensitive to the meanings of many function words (e.g., Lukatela, Crain, & Shankweiler, 1988).

[2] Linebarger et al. propose that the deficit resides in the mapping relation between the syntax and the semantics. They advance this "mapping hypothesis" to explain why their agrammatic subjects' performance on the grammaticality judgment task surpasses their performance on comprehension tasks like picture verification (see also Schwartz, Linebarger & Saffran, 1985). This hypothesis is extended in Schwartz, Linebarger, Saffran and Pate (1987) where the deficit is characterized as an inability to map syntactic functions onto appropriate thematic roles.

[3] An experiment by Foss (1969) presents an example of this phenomenon: it was found that response times are faster for detection of a target phoneme that comes late in a sentence, in comparison to one that comes early. It was Foss's view that this pattern required a structural explanation. Holmes and Forster (1970) reported the same pattern with a click-monitoring task, but they propose a different interpretation, attributing the longer reaction times for detecting early items to greater processing load at the early stages of sentence processing than at later stages. The inequality comes about putatively because at the beginning the subject is carrying out two tasks at the same time: making a decision about a target word and, in addition, processing the words that follow the target word. But if the detection of a target item is on-line, then it is unclear why varying the position of the target word should affect processing time in any systematic way.

The interference account also predicts that there should be no difference between detection tasks that present grammatically structured material and those that present unstructured word strings. But just this difference has been reported repeatedly in the literature. Marslen-Wilson and Tyler (1980) and Marslen-Wilson (1984) demonstrated that the word position effect is obtained with syntactically organized material, but it does not occur with syntactically scrambled word strings. Moreover, Aaronson (1968) reported that for presentation rates appropriate to conversational speech, response time to detect a target segment does not decrease but actually increases with serial position in a list of unrelated words. These findings suggest that the word position effect occurs only with syntactically organized material. Taken together, they lend considerable support to Foss's view that the word position effect reflects the processing of structure.

[4] For proposals of this kind to pay dividends it must turn out that agrammatism represents an underlying unity, theoretically and empirically. From the standpoint of theory, the notion of agrammatism must be characterized within a viable model of language performance; that is, it should conform to natural seams in the language apparatus. It is clear from the work of Grodzinsky (1984b), Kean (1980; 1982) and Lapointe (1983) that the phenomenon associated with agrammatism, the incorrect use of the closed-class morphology, satisfies this criterion.

Badecker and Caramazza (1985) question the status of agrammatism as an empirical entity in view of significant variations in language performance among aphasic individuals so designated. This is not the place for a discussion of the issues raised in their critique (see Caplan, 1986). We would point out that although we do not accept their conclusion that the concept of agrammatism should be jettisoned, we share a concern with the problems for classification that are created by intersubject variation, and we concur that the practice of grouping data over subjects should never be undertaken lightly.

[5] The proposal was extended by Grodzinsky (1984b) to take account of cross-language differences in the manifestations of agrammatism. In marking syntactic distinctions, Grodzinsky predicts misselection of closed-class items within the same syntactic category, with the proviso that closed-class items will usually be omitted in a relatively uninflected language like English, but not in a morphologically complex language like Hebrew, where failure to inflect the stem properly could result in a nonword.

[6] In addition, 32 simple conjoined-clause filler sentences were interspersed among the test items. These too consisted of equal numbers of grammatical and ungrammatical sentences.

[7] An idiosyncrasy of the stimulus set should be noted. Since it is impossible to construct sentences in which the position of the closed-class item is at the beginning of the sentence and also distant from the licensing word, there is unavoidably an imbalance in the data set. We have dealt with this circumstance in the following way: when we wished to analyze the effect of position as a factor, we deleted the sentences where the value for proximity was FAR. When we wished to look at proximity, we deleted the sentences where the value for position was BEGINNING. Accordingly, analyses of variance were carried out using the between-subjects factor of group, and within-subjects factors of position, *substitution type, proximity* and *category*.

[8] On the ungrammatical sentences, the mean response time was 2606 msec for the aphasic group and 1616 msec for the reference group. The difference was significant $F(1,10) = 15.62; p < .01$.

[9] The present study does not address the issue of the precise characterization of 'distance.' It may be that the amount of structure between the closed-class item and its licensing word is the relevant factor, or it may be the number of words, or amount of time. However distance comes to be defined, we will assume that our results are the effect of a single factor influencing the performance of both groups.

[10] An unexpected aspect of the ability to detect within-class substitutions is the detrimental effect on performance of the determiner substitutions. We do not have an adequate explanation of this effect. We do, however, consider it significant that the effect, whatever the cause, was evident for both groups. One might speculate that the increased response times result from the fact that the recognition of this type of error often depends upon a violation of number agreement between the determiner and its head noun. This type of error is often encountered in everyday contexts where the long latencies might reflect the time required to (unconsciously) correct this type of error (see Crain & Fodor, 1987).

[11] In a later paper, Grodzinsky (1984b) states the generalization as follows: if a terminal element at S-structure is not lexically specified (see Chomsky, 1981), then it will be unspecified at this level in an agrammatic's representation. In a reply, Sproat (1986) presents theoretical reasons against characterizing the agrammatic deficit as a structural deficit. He argues against the claim that in agrammatism traces are unspecified at S-structure. The argument has two parts. First, Sproat observes that the proposal entails that agrammatism involves the violation of a central tenet (the Projection Principle) of the linguistic framework in which Grodzinsky's hypothesis is placed (Government-Binding Theory). Sproat then points to evidence from the study by Linebarger et al. (1983a) showing that, in fact, the Projection Principle is respected in the grammars of the aphasics they tested.

[12] To explain the disparity between the aphasic subjects' ability to detect between- and within-class substitutions, we would suggest an alternative to Zurif and Grodzinsky's (1983) proposal. Our proposal invokes the concept of degrees of acceptability (Chomsky, 1965). Since a substitution that violates category membership constitutes a more egregious violation, it has greater perceptual salience than a violation that preserves category membership. On this account, it is not surprising that the subjects in both groups were quite accurate in identifying the more salient between-class substitutions (3% errors for the reference group: 12% for the aphasic group).

[13] These response times contrast sharply with post-sentential judgment tasks used to evaluate structural parsing preferences. For example, Frazier (1978) reports response times averaging over 1200 ms for the preferred reading of structural ambiguities. Although there are important differences between the stimulus materials in the present study and those used by Frazier, the differences in response times suffice to illustrate the point.

---

[14] Like Linebarger, Schwartz and Saffran (1983b), we find it implausible to suppose that the judgment task could permit syntactic structure to be computed by an alternative processing route. As these authors point out, this supposition amounts to the claim that there exists a separate, redundant sentence parser that is employed solely off line in certain tasks. It should also be mentioned that the Zurif and Grodzinsky proposal, which envisions such a delay between the arrival of an utterance and its structural analysis, entails unrealistic demands on working memory. That is, suppose that Broca's aphasics are unable to carry out on-line parsing. In order to judge grammaticality, then, they would be forced to accumulate and store a sentence as an unstructured aggregate of words that are subsequently fed into the special off-line parsing device. However, it seems to us that the verbal memory limitations of agrammatic aphasics would render them even less capable of such a feat than normal subjects.

Syntactic Comprehension in Young Poor Readers*

Suzanne T. Smith,[†] Paul Macaruso,[††] Donald Shankweiler,[††]
and Stephen Crain[††]

Children with specific reading disability fail to understand some complex spoken sentences as well as good readers. This investigation sought to identify the source of poor readers' comprehension difficulties. Second-grade good and poor readers were tested on spoken sentences with restrictive relative clauses in two experiments designed to minimize demands on working memory. The methodological innovations resulted in a high level of performance by both reader groups, demonstrating knowledge of relative clause structure. The poor readers' performance closely paralleled that of the good readers both in pattern of errors and in awareness of the pragmatic aspects of relative clauses. The findings suggest that limitations in processing account for comprehension difficulties displayed by some poor readers in previous investigations.

INTRODUCTION

This study investigates the syntactic abilities of poor readers, children who show a marked disparity between their measured level of reading ability and the level of performance that might be expected in view of their intelligence and opportunities for instruction. Much research on children with specific reading disability finds the source of their problems in the language domain, not in visual perception or general analytic ability (Liberman, 1983; Perfetti, 1985; Vellutino, 1979). Within the language domain, the research literature indicates that the difficulties poor readers display are usually associated with some aspects of phonological processing.[1]

Recently several investigations have raised the possibility that poor readers' limitations may include other components of language processing. For example, it has been found that poor readers fail to comprehend complex spoken sentences accurately in some circumstances. These findings have led some researchers to question whether poor readers have mastered all of the complex syntactic properties of the adult grammatical system.[2] However, it is important to appreciate that another explanation for their sentence processing difficulties can be given. It is possible that the problems poor readers have in comprehending sentences may ultimately derive from the same limitation in processing that accounts for their failures on lower-level tasks involving the phonology. The present research seeks to gain a better understanding of the basis of poor readers' failures in sentence comprehension by testing between these competing possibilities.

The first hypothesis, which we will call the *Syntactic Lag Hypothesis*, is a natural outgrowth of a large body of research on language acquisition that suggests that complex syntactic structures are late to develop. On this view, language is assumed to be acquired in a stage-like progression from simple to complex structures, with complex structures mastered later in the course of language acquisition (see Crain & Shankweiler, 1988, for further elaboration). Young poor readers are unable to comprehend certain syntactic structures as well as good readers because poor readers are considered to be developmentally delayed in the acquisition of syntactic knowledge (see e.g., Byrne, 1981; Fletcher, Satz, & Scholes, 1981; Stein, Cairns & Zurif, 1984). The Syntactic Lag Hypothesis is buttressed by findings indicating that, in samples of spoken language, poor readers tend not to use complex sentence structures, and their utterances often contain morphosyntactic errors (de Hirsch, Jansky, & Langford, 1966; Fry, Johnson, & Muehl, 1970; Morice & Slaghuis, 1985; Vogel, 1975). Further evidence shows that poor readers have difficulties with metalinguistic tasks such as correcting or explaining syntactically anomalous sentences (Bohannon, Warren-Leubecker, & Hepler, 1984; Bowey, 1986; Forrest-Pressley & Waller, 1984; Fowler, in press; Menyuk & Flood, 1981).

For an adherent to the Syntactic Lag Hypothesis, the reading comprehension failures of poor readers cannot be blamed on decoding difficulties alone, since on this account, poor readers have, in addition to their problems at the word level, a syntactic immaturity that predisposes them to difficulties in understanding some sentences. Moreover, the difficulties poor readers display in comprehending *spoken* sentences cannot be a result of their inadequacies in decoding. Thus, to account for poor readers' difficulties in reading and in comprehending spoken sentences with complex syntactic structures, proponents of the Syntactic Lag Hypothesis must invoke deficiencies in at least two components of the language system: the phonology and the syntax. On the one hand, a deficit in phonological processing must be invoked to account for weak decoding skills which are so often present in poor readers. On the other hand, their impaired sentence comprehension would reflect jointly the phonological processing problems and a delay in the acquisition of syntactic knowledge.

In contrast to this dual deficit account of poor readers' reading and comprehension difficulties, an alternative hypothesis explains both word-level and sentence-level problems as the result of the same underlying deficit. On this view, which we call the *Processing Limitation Hypothesis*, all of poor readers' language-related difficulties are considered to derive from a limitation in phonological processing. In other words, this hypothesis maintains that poor readers' performance deficits at higher levels of language processing are symptoms of an underlying phonological deficit (Liberman & Shankweiler, 1985; Mann, Shankweiler, & Smith, 1984; Perfetti, 1985; Shankweiler & Crain, 1986; Stanovich, 1986). In addition to creating difficulties in decoding, the phonological deficit gives rise to comprehension problems, both in reading and spoken language, by exhausting the resources of verbal working memory which must be shared by lower-level and higher-level language processes.

Evidence for the link between phonological processing and verbal working memory is based on research demonstrating that short-term memory for verbal material relies upon phonological coding (Conrad, 1964, 1972; Conrad & Hull, 1964; Crowder, 1978) and research indicating that poor readers show selective impairment on tests of verbal memory. Poor readers' impaired memory for verbal material has been demonstrated for a wide range of stimuli including nonsense syllables (Liberman, Mann, Shankweiler, & Werfelman, 1982), unrelated words (Mann, Liberman, & Shankweiler, 1980; Smith, Mann, & Shankweiler, 1986), and sentences (Mann et al., 1984; Mattis, French & Rapin, 1975; Roit, 1980; Shankweiler, Smith, & Mann, 1984; Weinstein & Rabinovitch, 1971; Wiig & Roach, 1975). In contrast, memory for a variety of nonverbal stimuli is unimpaired (Liberman et al., 1982; Nelson & Warrington, 1980; Vellutino, Pruzek, Steger, & Meshoulam, 1973).

With regard to sentence processing, the Processing Limitation Hypothesis supposes that because information in language is conveyed sequentially, phonological coding in verbal working memory is needed to retain linguistic information temporarily in verbatim phonological form while syntactic and semantic relations are determined. This approach advances the view that all the symptoms of reading disability arise from an underlying difficulty in phonological processing, thus eschewing the proposal that poor readers are delayed in the acquisition of syntactic knowledge. Instead, the Processing Limitation Hypothesis predicts that poor readers, because of their phonological processing deficit, will be unable to fully exploit their syntactic knowledge and will encounter difficulties with just those sentences that exceed the resources of verbal working memory. By contrast, on sentences that are less taxing of those resources, their performance level should approximate that of good readers.

To summarize, the Syntactic Lag Hypothesis and the Processing Limitation Hypothesis present sharply contrasting explanations of poor readers' problems in sentence comprehension. Since spoken language comprehension surely sets an upper bound on the ability to comprehend sentences in reading, it is imperative to reach a diagnosis of poor readers' comprehension problems. Neither of the hypotheses under consideration can be rejected at present. The aim of this research is to discover which one provides the best account of poor readers' problems in spoken-language comprehension. In the following experiments, the relative clause was chosen as a test case for evaluating the two hypotheses.

COMPREHENSION OF RELATIVE CLAUSES

Several studies of language acquisition have found that relative clauses emerge late in the course of normal language development (see Sheldon, 1974; Tavakolian, 1981). This has led proponents of both hypotheses to ask whether reading-disabled children are inferior in performance to age-matched good readers in comprehension of sentences containing these structures. Before we review studies that address this question, some preliminary comments are in order about the properties of relative clauses.

Relative clauses are of special interest because of their syntactic, and semantic/pragmatic properties. Four types of relative clause sentences are illustrated below. These are distinguished by a two-letter code indicating the noun phrase, subject or object, in the main clause that bears the relative clause and the site of the superficially empty noun phrase in the relative clause. For instance, in the OS sentence (3) the object noun phrase of the main clause is modified by a relative clause which lacks an overt noun phrase in subject position. The empty noun phrase is marked by "_" in these examples.

(1) SS The girl who _ pushed the boy tickled the clown.
(2) SO The boy who the girl pushed _ tickled the clown.
(3) OS The girl pushed the boy who _ tickled the clown.
(4) OO The boy pushed the girl who the clown tickled _.

In relative clause sentences, coreference must be established between the missing noun phrase in the relative clause and a controlling noun phrase in the main clause.

Coreference relations are governed by structural constraints involving the notion of c-command. These constraints rule out certain coreference possibilities and therefore restrict the semantic interpretations that a sentence can be assigned.

In addition, relative clauses serve several pragmatic functions. One pragmatic function distinguishes restrictive and nonrestrictive relative clauses. Restrictive relatives act to specify a particular subset from a larger set of items. This means that restrictive relatives are used felicitously in situations in which there is a group of objects corresponding to the head noun phrase of the relative clause. For example, in sentence (5a) the relative clause "that had a bow around its neck" acts to pick out a particular teddy bear from a group of teddy bears. In contrast, nonrestrictive relative clauses function merely to comment upon the noun phrases they modify, not to further delimit them. Sentence (5b) contains a nonrestrictive relative clause "which had a bow around its neck." In this case, the relative clause serves only as a further comment about the one teddy bear in the context, namely that it had a bow around its neck.

(5a) The boy was holding the teddy bear that had a bow around its neck.
(5b) The boy was holding the teddy bear, which had a bow around its neck.

Having reviewed the relevant properties of relative clauses, we turn now to studies of their comprehension by good and poor readers. Several recent investigations have indicated impaired performance by reading-disabled children on various forms of relative clause sentences (Byrne, 1981; Goldsmith, 1980; Mann et al., 1984; Stein et al., 1984). The findings have generally been regarded as supporting the Syntactic Lag Hypothesis. However, a dissenting note was sounded at the conclusion of the study by Mann et al. (1984), in which two of the present writers were co-investigators.

In the Mann et al. study, third-grade good and poor readers were tested for comprehension and (on another occasion) for repetition of the same relative clause sentences. On the comprehension test, the poor readers made more errors than the good readers; however, both reader groups made similar types of errors and there was no interaction between reader group and sentence type. On the repetition task, too, the poor readers performed less accurately than the good readers. Thus, the poor readers were impaired in both comprehension and repetition of test sentences, but their errors, although more numerous, were similar to those of the good readers. Mann et al. maintained that the higher error rate for the poor readers can be explained as stemming from their limitations in the effective use of phonological structure in the service of verbal working memory.

Giving credence to poor readers' limitations in processing, the present study sought to identify ways to reduce the processing load in order to permit them to demonstrate their knowledge of relative clauses. Until this step is taken, the source of poor readers' observed comprehension failures remains an open question; either the Syntactic Lag Hypothesis or the Processing Limitation Hypothesis could explain some portion of the findings. In planning the present study, several potential sources of processing difficulty that may have adversely affected poor readers' performance in earlier research were identified and steps were taken to minimize the difficulties. Success in reducing or eliminating the performance gap between good and poor readers would constitute support for the Processing Limitation Hypothesis. We now describe the factors that may affect difficulty of processing and the steps that were initiated to control these factors in the present study.

First, most of the previous studies of relative clauses failed to take into account their pragmatic functions. In other words, these studies did not meet the "felicity conditions" associated with the use of restrictive relative clauses. For example, the Mann et al. study used an object manipulation test in which the stimulus array consisted of one token of an object for each noun phrase. As discussed earlier, restrictive relative clauses are used felicitously only in contexts which establish a set of objects corresponding to the head noun phrase of the relative clause. To satisfy this pragmatic condition, more than a single item corresponding to the head noun phrase of the relative clause should be made available in the experimental workspace. Evidence from recent studies of language acquisition has demonstrated the importance of satisfying the felicity conditions associated with the presuppositions of restrictive relative clauses and the detrimental effect that failure to meet these conditions has upon young children's comprehension. For example, Hamburger and Crain (1982) found significant improvement in preschool children's comprehension of relative clause sentences when felicity conditions were satisfied. They present evidence that improved performance derives from a reduction in processing demands associated with the planning required to make a response in an act out task. It is possible that poor readers' difficulties with relative clause sentences might also be tied to excessive demands imposed by failure to meet felicity conditions.

Another finding from the literature on language acquisition suggests that poor readers' impaired performance in the Mann et al. study may have been due, in part, to the number of relations among animate noun phrases that needed to be determined and acted out. In a study with preschool children, Goodluck and Tavakolian (1982) demonstrated that performance on an act-out task similar to the one used in the Mann et al. study improved significantly when the number of animate noun phrases in relative clause sentences was reduced from three to two (e.g., by replacing a transitive verb in the relative clause with an intransitive verb). Improved performance in this case may also be explained in terms of a reduction in processing demands associated with the response plan needed to act out these sentences (Hamburger & Crain, 1984).

Finally, it was observed by Goldsmith (1980) that poor readers' comprehension of semantically constrained sentences surpassed that of unconstrained sentences. Semantically anomalous relations which have often characterized sentences used in earlier studies (e.g., Byrne, 1981; Mann et al., 1984) may require additional processing in order to recover the intended relations among the constituents. Poor readers, with special weaknesses in verbal working memory, are presumably at a particular disadvantage under any circumstances that require reanalysis (Shankweiler et al., 1984).

Since only performance, not competence, is observed in tests of language ability, failure to control for factors that can unduly tax limited processing resources can mask expression of underlying linguistic knowledge. We suspect that previous studies of poor readers' comprehension of complex syntactic structures have often resulted in underestimates of their linguistic knowledge because, unwittingly, the experimental tasks posed extraneous demands on processing. When the goal of the research is to draw inferences about linguistic competence, experiments with the potential for disentangling deficits in structural knowledge from processing limitations are required. Bearing in mind the potential benefit of reducing processing demands for poor readers' comprehension of relative clauses, we incorporated four methodological changes to minimize processing load: (i) the restrictive function associated with restrictive relative clauses was satisfied; (ii) each of the test sentences indicated a single action relating two animate noun phrases; (iii) in each test sentence, a third noun phrase was inanimate and conveyed descriptive information; (iv) all test sentences indicated plausible events. Together these steps were expected to minimize the demands associated with the comprehension task and permit the subjects to display their full level of linguistic competence.

This investigation was designed to permit a direct comparison of the findings with those of the Mann et al. study. By comparing performance in the two studies, we can begin to evaluate evidence for the two hypotheses under consideration. If the Syntactic

Lag Hypothesis is correct, we would expect that poor readers should show no improvement when the task is made easier by minimizing processing demands because in the absence of requisite structures, impaired performance should be observed even under the simplest conditions. On the other hand, if the Processing Limitation Hypothesis is correct, we would expect poor readers' performance to improve as the task is simplified, reducing the disparity with good readers.

EXPERIMENT I: OBJECT MANIPULATION TEST

An object manipulation test was constructed in accordance with the methodological prescriptions described in the preceding paragraphs. This test incorporates the same structures studied in the research of Mann et al. and is intended as a sequel to that investigation.

Method

Subjects

The subjects were selected from four second-grade classes in a suburban public school system. All children for whom parental consent was obtained were given the Decoding Skills Test (DST) (Richardson and DiBenedetto, 1986) as a test of reading ability and the Peabody Picture Vocabulary Test—Revised (PPVT) (Dunn and Dunn, 1981). The PPVT was included to ensure that any obtained reader group differences could not be attributed to limitations in word knowledge, which is an indicator of general verbal ability.

Nonoverlapping groups of good and poor readers were formed on the basis of three criteria: a combined word and nonword DST score less than 64 (poor readers) or greater than 83 (good readers),[3] teacher confirmation of reading ability and a PPVT score between 90 and 130. Grouping by these criteria yielded 18 children in the poor reader group (8 boys; 10 girls) and 16 children in the good reader group (6 boys; 10 girls). The groups did not differ significantly in age, (range: 90 to 104 months), nor in PPVT scores (range: 90 to 126).[4] Additionally, all subjects were native speakers of English and had no known speech or hearing deficiencies. The mean age, PPVT score, and DST score for each group are summarized in Table 1.

TABLE 1. Characteristics of the subjects.

	Good Readers		Poor Readers	
	(N = 16)		(N = 18)	
	Mean	S.D.	Mean	S.D.
Age in Months	97.75	3.64	97.05	4.77
PPVT-Revised Score	114.93	9.86	110.00	8.53
DST Score*	98.37	8.63	47.77	15.53

*Maximum DST score = 120

Materials

A. Test Sentences: Four types of relative clause sentences were included in the test materials. An example of each is given below:

(6) SS The lady who held an umbrella kissed the man.
(7) SO The man who the lady kissed held an umbrella.
(8) OS The lady kissed the man who held an umbrella.
(9) OO The lady kissed the man who an umbrella covered.

Each sentence contains two animate noun phrases and one inanimate noun phrase and denotes plausible events among highly recognizable objects. The relations between animate noun phrases are reversible. In order to preserve the plausibility requirement, an action verb is used in the relative clause in OO-type sentences. For SO sentences, the inanimate noun phrase is placed in the main clause instead of in the relative clause to preserve plausibility and to allow for the possibility of a conjoined-clause misanalysis.[5]

Conjoined-clause sentences (CC), derived from the OS prototypes, were also included. It has been claimed that these structures are syntactically less complex than relative clause sentences and are mastered earlier in development. Each CC sentence contains an empty noun phrase in the second clause that is coreferential with the subject of the first clause. A sample CC is given in (10):

(10) CC The lady kissed the man and _ held an umbrella.

B. Test Design: The test consisted of five sets of sentences in which the relations among nouns were varied in the five sentence types (SS, SO, OS, OO, CC). This yielded five sentences of each type, making a total of 25. These were randomized and two test orders were prepared. Each test order was preceded by a short pretest which consisted of three simple sentences and one complex sentence. The pretest and test sentences were recorded on audiotape at a natural conversational rate. Each sentence was preceded by an alerting bell. The pretest and test sentences are listed in Appendix A.

Procedure

The subjects were tested individually in three sessions during the last quarter of the school year. The DST and PPVT were administered in the first session and one of the experimental tests was administered in each of the remaining sessions. Half of each reader group was tested on the Object Manipulation Test in the second session and half received the Sentence-Picture Matching Test. This order was reversed in the third session, thus counterbalancing the presentation of the two tests.

At the onset of the Object Manipulation Test, the subject was shown all of the toy objects to be used in the task and asked to name them. Any name that did not correspond to the name used in the test sentences was corrected. Before each test sentence was presented, objects corresponding to the nouns in the sentence were placed on the tabletop in front of the subject in a fixed, random order. As indicated earlier, two objects corresponding to the head noun in the relative clause were presented to meet the normal presuppositions associated with use of restrictive relative clauses. Although there are no such presuppositions associated with conjoined-clause sentences, an extra animate agent was placed in the work space to keep constant the number of items across sentence types. The subject listened to the test sentences using earphones at sound levels adjusted for comfortable listening.

The subject was instructed to listen to the entire sentence before proceeding to act out its meaning. Responses were coded by the experimenter to indicate the actions and relations enacted, as well as their order of occurrence. A sentence was repeated a second time if the subject requested it.

Results

Separate analyses were performed on the relative clause sentences and the conjoined-clause sentences in order to permit comparison with the results of the Mann et al. (1984) study. A response was considered correct if the relations indicated in both the main clause and the relative clause were properly enacted regardless of which clause was acted out first. Since there were no effects of test order, the remaining analyses were performed on the combined test orders.

Relative Clause Sentences

Both the good and poor readers acted out the relative clause sentences with a high level of success. The reader groups did not differ significantly in mean percentage of errors (good readers: 11.3%; poor readers: 18.1%). Table 2 displays the mean percentage of errors and standard deviations for each type of relative clause sentence for both reader groups.

TABLE 2. Object Manipulation Test.

	Mean percentage of errors and standard deviations for relative clause sentences.			
Relative Clause Structure	Good Readers		Poor Readers	
	Mean	S.D.	Mean	S.D.
SS	2.6	6.8	3.4	7.6
SO	25.0	28.8	35.6	26.2
OS	11.2	14.6	20.0	22.8
OO	6.2	12.0	13.2	21.8

The main effect of sentence type (SS, SO, OS, and OO relative clause sentences) was highly significant, $F(3,96) = 17.94$, $p < .001$. There was, however, no significant interaction between reader group and sentence type, indicating that the good and poor readers were similarly affected by variations in sentence structure. Scheffe post hoc analyses ($p = .05$) indicated that SO sentences were more difficult than OO and SS sentences, but were not more difficult than OS sentences. The SS sentences did not differ from OO sentences, but did differ from OS and SO sentences. Finally, OS and OO sentences did not differ from each other.

Analysis of errors. The distribution of errors was examined to determine whether errors previously identified in the literature as possibly signifying an immature stage of syntactic development occurred more often for the poor readers than for the good readers. Although the incidence of conjoined-clause errors differed significantly across sentence types, $F(2,64) = 3.52$, $p = .03$ (with SO sentences producing the most errors of this type), there was no main effect of reader group and no reader group x sentence type interaction. The good and poor readers did not differ in the number of conjoined-clause errors, and the pattern of errors across sentence type was similar for both reader groups. Also, other error types noted in previous investigations, e.g., minimum distance errors[6] that were noted in the Mann et al. (1984) study, failed to distinguish between reader groups in the present study.

There was no difference between reader groups in the order in which the information in the two clauses was acted out. In general, both reader groups acted out the descriptive information first. For example, in the SO sentence (7) "The man who the lady kissed *held an umbrella*," both the good and poor readers put the umbrella in the man's hand before they had the lady kiss him.[7]

Conjoined-Clause Sentences

The good readers made significantly fewer errors (2.6%) than the poor readers (17.8%) on the conjoined-clause sentences $t\,(32) = 2.71$, $p = .01$. Despite the poor readers' greater percentage of errors on this sentence type, the order in which they acted out the clauses did not differ significantly from that of the good readers. Enactments in which the second-mentioned clause was acted out first occurred 67.6% and 78.8% of the time for the good and poor readers, respectively.

Discussion

As noted in the Introduction, an examination of both error rates and pattern of errors by good and poor readers is critical for evaluating the two hypothesized sources of poor readers' difficulties in comprehension. We are justified in concluding that the Syntactic Lag Hypothesis is correct only if poor readers' percentage of errors, types of errors and items on which they err are discrepant from that of good readers. The Processing Limitation Hypothesis is correct only if the following criteria are met: (i) both reader groups are similarly affected by syntactic complexity—that is, those relative clause sentences that are most difficult for poor readers are similarly the most difficult for good readers; (ii) the types of errors made by poor readers are not qualitatively different from those made by good readers; (iii) there is improved performance on tests in which processing demands have been reduced as compared to tests in which they were not reduced; (iv) the performance of poor readers is well above chance on some examples of the structure being investigated.

The findings of this experiment clearly support the Processing Limitation Hypothesis. The following facts are indicative: First, there was no difference in the incidence of errors between the reader groups on relative clause sentences. Second, both reader groups were similarly affected by syntactic variation as demonstrated by the lack of a reader group by sentence type interaction. Finally, the types of errors that the poor readers made were similar to those made by the good readers, and no error type was made more often by the poor readers. These findings are inconsistent with the Syntactic Lag Hypothesis. Instead, they suggest that the Processing Limitation Hypothesis may provide a better account of poor readers' inferior performance in earlier studies.

The effects of the steps we adopted to ease processing demands can be evaluated by comparing the good and poor readers' performance in this study with the reader groups' performance in the Mann et al. (1984) study in which animacy, plausibility, and felicity conditions were not controlled. Figure 1 presents a comparison of the percentage of errors for the good and poor readers on the Object Manipulation Test in the two studies.

Both reader groups in the present study performed at a significantly higher level than those in the Mann et al. study (even though the subjects were a year younger in the present study). The good readers made 11.3% errors compared with 25.7% in the Mann et al. study $t\,(32) = 2.81$, $p < .01$, and the poor readers made 18.1% errors compared with 40.8% in the earlier study $t\,(33) = 4.01$, $p < .001$. However, it should be noted that the same pattern of errors across sentence types was found in both studies. Also, in contrast to the Mann et al. study, there were no differences between the reader groups in this study for any of the prevalent error types. In sum, the reduction in percentage of errors that occurred in the present study strongly suggests that poor readers' impaired performance in earlier studies is attributable to processing factors, and not to a delay in the acquisition of syntactic knowledge.

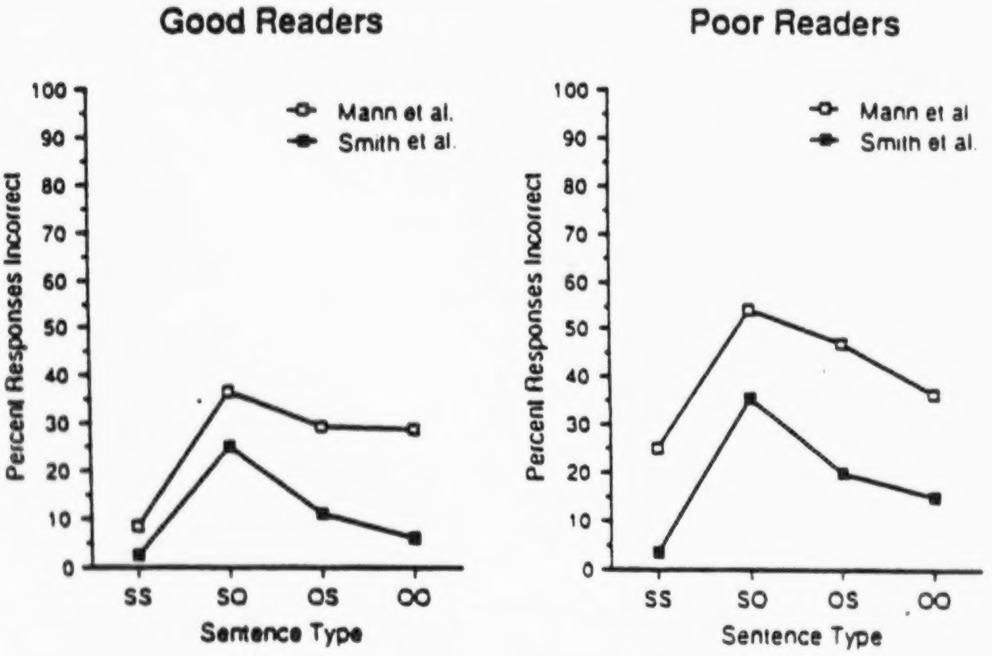


Figure 1. Comparison of the percentage of errors for the good and poor readers on the Object Manipulation Test in this study and in the Mann et al. (1984) study.

A further aspect of the poor readers' performance on the Object Manipulation Test supports this conclusion. This concerns the unexpectedly high incidence of errors on the conjoined-clause sentences. The difficulty, we suspect, may have been due to the unusual construction of these sentences. As explained earlier, these sentences were derived from the OS relative clause sentences in order to investigate the possibility that some subjects were adopting the conjoined-clause analysis posited by Tavakolian (1981). In the conjoined-clause test sentences, the descriptive information specified in the OS relative clause appears *after* the overt action is specified, as in (11), creating a conflict between the conceptual order of the information and its order of mention.

(11) CC The lady kissed the man and held an umbrella.

Intuitively, one might suppose that in normal conversation the information in this sentence would have been conveyed in the following manner:

(12) The lady held an umbrella and kissed the man.

This intuition is supported by the finding that the second clause was enacted first in 72.3% of all responses. The subjects appear to have treated the information in the second clause as if it were descriptive information representing some prior condition, even though conjoined-clause sentences do not have the inherent ordering of clauses characteristic of sentences with restrictive relative clauses. Given that these conjoined-clause sentences conveyed information in a way that may have unintentionally violated conversational conventions, they may have imposed a heavy processing load. Consequently, the poor readers were adversely affected.[8] This finding underscores our contention that poor readers' performance on tests of comprehension is closely tied to aspects of processing, and not necessarily to aspects of syntactic complexity.

In summary, the main finding was that the poor readers' performance on the relative clause sentences in the Object Manipulation Test did not differ from the good readers' performance either in percentage of errors or in pattern of responses. Also, the incidence of errors was significantly reduced in comparison with the Mann et al. study, even though the subjects were a year younger. Taken together, these findings support the Processing Limitation Hypothesis over the Syntactic Lag Hypothesis.

EXPERIMENT II: SENTENCE-PICTURE MATCHING TEST

The second experiment was designed to test further for the possibility that certain sentences containing relative clauses are misconstrued by poor readers as sentences containing conjoined clauses. This type of error in comprehension occurred relatively frequently in some previous research on interpretations of relative clause sentences by reading disabled children (Mann et al., 1984; Stein et al., 1984). It was reasoned that if the Syntactic Lag Hypothesis is correct, we would expect poor readers to be more susceptible to a conjoined-clause interpretation than good readers. The Processing Limitation Hypothesis, on the other hand, would expect no qualitative differences between reader groups.

A sensitive test of a tendency to construe sentences containing a relative clause as though they contain conjoined clauses can be made by forcing a choice between two pictures, one of which depicts the conjoined-clause analysis. (This technique was used successfully in examining comprehension of relative clauses by much younger children; see Crain and Fodor, forthcoming). Accordingly, a two-choice Sentence-Picture Matching Test was constructed by modifying the sentence material in Experiment I and providing appropriate pictures. For each SO, OS, and OO relative clause sentence one picture foil depicted the incorrect conjoined-clause analysis interpretation. Since a conjoined-clause analysis interpretation is equivalent to the correct interpretation of an SS relative clause sentence, comprehension of this sentence type was examined by contrasting the correct interpretation with a foil which depicted the interpretation of the information in the main clause only.

Method

Materials

A. Test Sentences: The five types of stimulus sentences used in Experiment I were modified in this experiment by changing the verb tense from past to present progressive to accommodate the use of a two-choice Sentence-Picture Matching Test. A sample set of test sentences is listed below:

(13) SS The lady who is holding an umbrella is kissing the man.
(14) SO The man who the lady is kissing is holding an umbrella.
(15) OS The lady is kissing the man who is holding an umbrella.
(16) OO The lady is kissing the man who an umbrella is covering.
(17) CC The lady is kissing the man and is holding an umbrella.

Additional sentences, with different nouns and verbs, were added as controls to ascertain that the subjects were attending to the entire sentence. These control sentences were of the same form as three of the experimental sentences (SS, OS, CC), but their respective foils differed. The picture foils for all sentence types are described in section C below.

B. Test Design: The test contained 40 sentences: the five (modified) sets of sentences (SS, SO, OS, OO, CC) from the Object Manipulation Test plus five sets of the three control sentences (SS, OS, CC). Two test orders were prepared, corresponding to those in the Object Manipulation Test, but with control sentences interspersed. A pretest (modified from Experiment I) was provided. All the items are listed in Appendix B. The test sentences were recorded on audiotape at a natural speaking rate by the same speaker who recorded the sentences in Experiment I. Each sentence was preceded by an alerting bell.

C. Picture Foils: The notation introduced by Sheldon (1974) is used here to describe the foils for sentences in the Sentence-Picture Matching Test. In this notation, noun phrases are numbered consecutively in the order in which they appear in the sentence. The subject and object relations of each verb phrase are indicated by a pair of numbers. For example, the correct interpretation of an SS sentence is 12,13. Thus, the correct response to sentence (18) has the woman (1) holding an umbrella (2) and kissing a man (3).

1 2 3

(18) SS The lady who is holding an umbrella is kissing the man.

As we have seen, there are two possible conjoined-clause analyses for SO sentences (12,13 and 21,23). One of these (12,13) was the most commonly observed conjoined-clause response in both the Tavakolian (1981) and the Mann et al. (1984) studies. It was therefore selected as the foil for this sentence type. This analysis of sentence (14) results in the interpretation, "the man is kissing the lady and is holding an umbrella."

For OS sentences, only one conjoined-clause interpretation (12,13) is possible. For sentence (15) this interpretation yields "the lady is kissing the man and is holding the umbrella." The OO sentences, like the SO, offer two conjoined-clause analyses (12,13 and 12,31). Young children in the Tavakolian study gave one of these responses (12,31) slightly more often than the other. This was also the predominant error response to OO sentences in the Mann et al. study. Therefore, this error was incorporated into the foil for the OO sentence type. For sentence (16) the foil picture had the lady kissing the man and the umbrella covering the lady.

A foil depicting a main-clause only interpretation was used for the target SS sentences. For sentence (13) this yields an interpretation of "the lady is kissing the man." For the conjoined-clause sentences, an erroneous minimum-distance principle interpretation (12,13) was selected as a foil. For example, in sentence (17), this shows the lady kissing a man and the man holding an umbrella. For the control SS and OS sentences, a relative-clause-only interpretation was depicted in the foil (12 and 23, respectively). Finally, a first-clause-only (12) interpretation was used for the control CC sentences.

D. Picture Specifications: In order to satisfy the felicity conditions associated with restrictive relative clauses (Hamburger & Crain, 1982), two tokens of the object corresponding to the empty noun phrase in the relative clause were presented in each correct picture. The additional objects were placed in the background. For the foils, an extra object which did not satisfy the felicity condition of the relative clause, but which was appropriate for the sentence context, was included in the background to maintain the same number of objects in each picture.

Each picture measured 4 x 6 inches, and was displayed on an 8 1/2 x 11 inch sheet with a one-inch border on the top, bottom and right side, and a 1 1/2 inch border on the left side to accommodate binding. In half of the arrays, the correct picture was placed at the bottom. A sample OS picture array is displayed in Appendix C.

Procedure

Administration of the Sentence-Picture Matching Test was counterbalanced with the Object Manipulation Test (see Experiment I: Method). The subjects, each of whom also participated in Experiment I, were tested individually. For each trial, the picture array was presented immediately before the onset of the tape-recorded sentence. The subjects were instructed to listen carefully to the entire sentence, to look carefully at both pictures in the array, and then to point to the picture that showed the meaning of the sentence. Headphones were used with the sound level adjusted for each subject's comfort. A sentence was repeated a second time if the subject requested it. .

Results

Separate analyses were performed on the relative clause sentences, the conjoined-clause sentences, and the control sentences. Since there was no effect of test order, the remaining analyses were performed on the combined test orders.

Table 3 displays the mean percentage of errors and standard deviations for each type of relative clause sentence for both reader groups. The good readers averaged 11.3% errors and the poor readers 18.6% errors, $F(1,32) = 4.69$, $p = .04$. Although the poor readers chose the erroneous conjoined-clause interpretation of relative clause sentences slightly (but significantly) more often than the good readers, both reader groups achieved a high overall level of success. In fact, the difference between the mean percentage of errors for the two groups on this test was very similar to the difference obtained with the Object Manipulation Test. Further, as was the case with the Object Manipulation Test, there was a highly significant effect of sentence type, $F(3,96) = 21.89$, $p < .001$, but no reader group x sentence type interaction. Post hoc Scheffe analyses ($p = .05$) indicated that SO sentences were more difficult than all other sentence types. The OO and OS sentence types were not different from each other, but both were more difficult than SS sentences.

TABLE 3. Sentence-Picture Matching Test.

Mean percentage of errors and standard deviations for relative clause sentences.

Relative Clause Structure	Good Readers		Poor Readers	
	Mean	S.D.	Mean	S.D.
SS	0.0	–	1.2	4.8
SO	22.6	17.4	38.8	31.8
OS	5.0	11.4	14.4	17.8
OO	12.6	17.8	20.0	19.4

There was no significant difference between reader groups in the percentage of errors on conjoined-clause sentences (good readers: 7.6%; poor readers: 3.4%). Only a single error occurred on the control sentences.

Discussion

On the Sentence-Picture Matching Test, there was a small, but significant difference between the reader groups in the mean percentage of errors, i.e., incorrect conjoined-clause interpretations, for the relative clause sentences. However, no subjects consistently chose this interpretation across all sentence types. Further, the difference between the mean percentages of errors for the two groups on the Sentence-Picture Matching Test was very similar to the difference found on the Object Manipulation Test, and both reader groups were similarly affected by variations in syntactic structure introduced by the different types of relative clause sentences. The fact that the poor readers gave a high percentage of correct responses on the Sentence-Picture Matching Test (and on the Object Manipulation Test), together with the fact that they showed a similar pattern of errors across sentence types as the good readers in both tasks, suggests that the difficulties poor readers have shown with relative clause sentences in past studies cannot be attributed to a syntactic lag.

In contrast with the results from the Object Manipulation Test, in which the reader groups differed only on conjoined-clause sentences, in the Sentence-Picture Matching Test conjoined-clause sentences yielded no reader group difference. We can only speculate about the reasons for this disparity. Certainly, the fact that the poor readers did not differ from the good readers in the percentage of errors on conjoined-clause sentences in this experiment is in keeping with the lower overall processing demands imposed by a forced-choice selection procedure (see Shankweiler et al., 1984). In this case, the subjects merely had to affirm or deny the accuracy of what was depicted, rather than plan and execute a set of actions demonstrating comprehension of this information. This interpretation gains cogency in light of the analysis by Hamburger and Crain (1984), who have explicitly characterized the processing demands associated with the planning step in act-out tasks. They demonstrated, moreover, that by minimizing the processing demands associated with the planning step, young children's performance is enhanced. The finding that the poor readers were not impaired in comprehension of conjoined-clause sentences when the processing load was minimized is, of course, consistent with the Processing Limitation Hypothesis. This finding cannot easily be assimilated by the Syntactic Lag Hypothesis.

GENERAL DISCUSSION

The aim of this research was to pinpoint the source of poor readers' comprehension failures on spoken sentences containing complex syntactic structures. Sentences with relative clauses provided a vehicle for testing between two hypotheses either of which could, in principle, explain the comprehension difficulties uncovered in previous research. As we saw, the Syntactic Lag Hypothesis views poor readers as delayed in the acquisition of complex syntactic structures. Deficits in comprehension thus reflect a lack of full linguistic competence on the part of poor readers. On this account, poor readers are expected to encounter difficulty with putatively late-developing complex syntactic structures such as relative clauses. Their errors, too, should reflect their relatively immature grammars. For instance, poor readers might be expected to make more conjoined-clause analysis errors than good readers if adult-like recursive rules have yet to become a part of their grammatical knowledge. Further, since the requisite structures are hypothesized not to be in place, poor readers should be expected to make more errors than good readers in spite of steps one might take to reduce processing demands.

By contrast, the Processing Limitation Hypothesis maintains that even the most complex syntactic structures are in place at the onset of reading instruction. Errors in comprehension on the part of poor readers are not the result of gaps in linguistic competence, but instead, are taken to reflect performance failures deriving from deficiencies in verbal working memory due to limitations in phonological processing. The Processing Limitation Hypothesis, but not the Structural Lag Hypothesis, anticipated that poor readers would achieve a high level of performance on the tasks employed in the present study, which reduce the burdens on verbal working memory. Further, the Processing Limitation Hypothesis anticipated a substantial reduction in errors in the present study as compared to previous studies in which steps were not taken to hold processing demands to a minimum. Finally, this hypothesis predicted that poor readers' overall performance should parallel that of good readers.

In keeping with these predictions, the poor readers in the present study performed at a high level of success on both experimental tasks. On the Object Manipulation Test, they were not differentiated from the good readers either with respect to percentage of errors or types of errors. Further, there was a significant reduction in the percentage of errors as compared to the object manipulation test of Mann et al. (1984), a difference that must reflect the effectiveness of our attempt to simplify the task. Finally, the poor readers responded in the same conceptual manner as the good readers. On the Sentence-Picture Matching Test, there was a marginally significant reader group difference. However, in neither experiment did the poor readers deviate from the good readers on any measure that could be taken to indicate delayed acquisition of syntactic knowledge of relative clauses, e.g., different types of errors, or a different rank order of difficulty across types of relative clause sentences. Taken together, these findings support the expectation of the Processing Limitation Hypothesis that both reader groups are indistinguishable in syntactic competence, and that any differences in comprehension in children differing in reading ability must be attributable to other causes. We should not lose sight of the fact that the subjects in the present study were a year younger than those tested in the Mann et al. study. This adds to the weight of the evidence in favor of the Processing Limitation Hypothesis. Under the Syntactic Lag Hypothesis, younger poor readers would more likely show evidence of delayed acquisition of complex syntactic structures. But since the younger children in the present study were able to comprehend sentences containing complex relative clause constructions when processing demands were held to a minimum, the Syntactic Lag Hypothesis is rendered less plausible as an explanation of poor readers' comprehension difficulties.

Earlier, we noted that findings over a range of measures indicate that poor readers are deficient in phonological processing. It has been hypothesized that one consequence of such a deficit is failure to apprehend the phonological structure of words, making it inordinately difficult for the child to discover how the spelling of the word is related to its phonological structure. Thus, on this hypothesis, the development of efficient word recognition strategies is hampered because the affected individual has not learned to decode. As indicated earlier, there exists considerable empirical support for this claim. But decoding difficulties are not the only consequence of a deficit in phonological processing: a phonological deficit may also impose severe limitations on the operation of verbal working memory. Sentence comprehension may be affected adversely both in reading and in the apprehension of spoken language.

Let us now spell out how a phonological processing deficit may result in failures in sentence comprehension through its impact on verbal working memory. On this account, which is presented in greater detail elsewhere (see Crain et al., in press; Shankweiler & Crain, 1986), the well attested deficiencies of poor readers on tasks of ordered recall lead to the expectation that higher level processes such as syntactic comprehension would be compromised by poor readers' failures to adequately retain phonological information during sentence processing. In effect, there is a "bottleneck" that constricts the flow of information from lower levels to higher levels (see Perfetti, 1985, for a related account). This follows from a very general assumption concerning the architecture of the language apparatus. To explain how a deficit at one level can give rise to dysfunction at other levels, we have appealed to the proposition that the language apparatus consists of hierarchically organized subcomponents with information flowing unidirectionally from lower to higher levels of the system.[9] A role for verbal working memory is posited to support processing within each structural component of language and to effect the orderly transfer of information between components. (Other accounts which emphasize the control functions of working memory in addition to its short-term storage functions have been advanced by Baddeley & Hitch, 1974, and Daneman & Carpenter, 1980).

This conception of poor readers' difficulties in processing spoken sentences applies with even greater force to reading, because reading imposes an additional step upon the language understanding system—a phonological interpretation must be assigned to an orthographic representation by a decoding step. Whereas in spoken language word recognition is carried out quite automatically by machinery which evolution has crafted for this purpose, reading depends on orthographic skills that are not part of the biological endowment for language (see Liberman, 1988, for discussion). When orthographic decoding skills are poorly established, as in the case of a beginning reader or a person with a persistent reading disability, syntactic and semantic processing is severely hampered because the limited verbal working memory resources are used up in attempts to identify the words of the text. Thus, until word recognition skills are well established, reading may be expected to impose greater demands on processing capacities than speech (Shankweiler & Crain, 1986). Here, as in the case of spoken sentence comprehension, failures to understand complex structures may be attributable to limitations in phonological processing and not to a delay in the acquisition of syntactic structures. Therefore, in reading comprehension, as in spoken language comprehension, we should observe improved performance in higher level processes when demands associated with lower level components of the task are reduced.

APPENDIX A

Stimulus Sentences for Object Manipulation Test

SS	The lady who held an umbrella kissed the man.
SO	The man who the lady kissed held an umbrella.
OS	The lady kissed the man who held an umbrella.
OO	The lady kissed the man who an umbrella covered.
CC	The lady kissed the man and held an umbrella.
SS	The lady who carried a suitcase touched the man.
SO	The lady who the man touched carried a suitcase.
OS	The man touched the lady who carried a suitcase.
OO	The man touched the lady who a scarf was on.
CC	The man touched the lady and carried a suitcase.
SS	The boy who carried a banana chased the monkey.
SO	The boy who the monkey chased carried a banana.
OS	The monkey chased the boy who carried a banana.
OO	The monkey chased the boy who a hat was on.
CC	The monkey chased the boy and carried a banana.
SS	The girl who hugged a teddy bear pushed the boy.
SO	The boy who the girl pushed hugged a teddy bear.
OS	The girl pushed the boy who hugged a teddy bear.
OO	The boy pushed the girl who the ice cream fell on.
CC	The boy pushed the girl and hugged a teddy bear.
SS	The giraffe that ate the hay followed the elephant.
SO	The giraffe that the elephant followed ate the hay.
OS	The elephant followed the giraffe that ate the hay.
OO	The elephant followed the giraffe that a rope hung from.
CC	The giraffe followed the elephant and ate the hay.

Pretest

The girl pushed the boy.
The man touched the lady.
The monkey followed the giraffe.
The boy held a teddy bear and the girl carried a ball.

APPENDIX B

Stimulus Sentences for Sentence-Picture Matching Test

SS	The lady who is holding an umbrella is kissing the man.
SO	The man who the lady is kissing is holding an umbrella.
OS	The lady is kissing the man who is holding an umbrella.
OO	The lady is kissing the man who an umbrella is covering.
CC	The lady is kissing the man and is holding an umbrella.

SS The lady who is carrying a suitcase touched the man.
SO The lady who the man is touching is carrying a suitcase.
OS The man is touching the lady who is carrying a suitcase.
OO The man is touching the lady who a scarf is on.
CC The man is touching the lady and is carrying a suitcase.

SS The boy who is carrying a banana is chasing the monkey.
SO The boy who the monkey is chasing is carrying a banana.
OS The monkey is chasing the boy who is carrying a banana.
OO The monkey is chasing the boy who a hat is on.
CC The monkey is chasing the boy and is carrying a banana.

SS The girl who is hugging a teddy bear is pushing the boy.
SO The boy who the girl is pushing is hugging a teddy bear.
OS The girl is pushing the boy who is hugging a teddy bear.
OO The boy is pushing the girl who the ice cream is falling on.
CC The boy is pushing the girl and is hugging a teddy bear.

SS The giraffe that is eating the hay is following the elephant.
SO The giraffe that the elephant is following is eating the hay.
OS The elephant is following the giraffe that is eating the hay.
OO The elephant is following the giraffe that a rope is hanging from.
CC The giraffe is following the elephant and is eating the hay.

Control Sentences

SS The lady who is wearing a hat is hugging the baby.
OS The lady is hugging the baby who is wearing a hat.
CC The lady is hugging the baby and is wearing a hat.

SS The clown who is holding the flowers is hitting the girl.
OS The girl is hitting the clown who is holding the flowers.
CC The clown is hitting the girl and is holding the flowers.

SS The girl who is holding a ball is kicking the boy.
OS The boy is kicking the girl who is holding a ball.
CC The boy is kicking the girl and is holding a ball.

SS The boy who is eating the ice cream is pulling the girl.
OS The boy is pulling the girl who is eating the ice cream.
CC The girl is pulling the boy and is eating the ice cream.

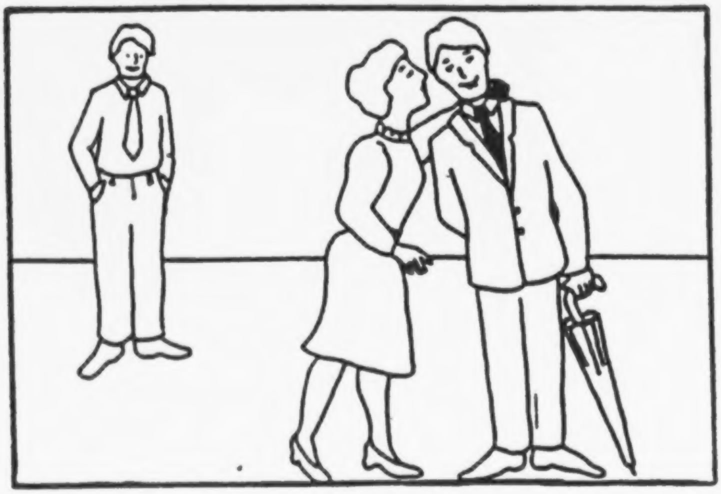
SS The cat that is wearing a bow is licking the rabbit.
OS The rabbit is licking the cat that is wearing a bow.
OO The cat is licking the rabbit and is wearing a bow.

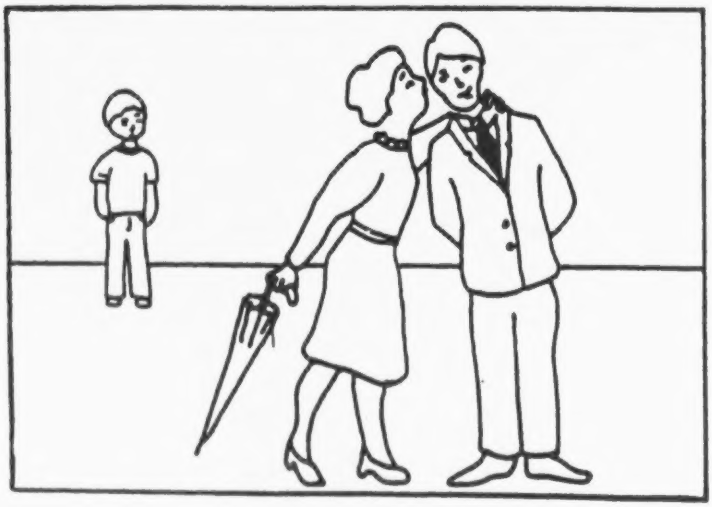
Pretest

The girl is pushing the boy.
The man is touching the lady.
The monkey is following the giraffe.
The boy is holding a teddy bear and the girl is carrying a ball.

APPENDIX C

Sample Picture Array for Sentence-Picture Matching Test





OS The lady is kissing the man who is holding an umbrella.

ACKNOWLEDGMENT

This paper is based on a doctoral dissertation by the senior author. The research and preparation of the manuscript were supported in part by a Doctoral Dissertation Fellowship from the University of Connecticut and a Program Project Grant to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994). We wish to thank the students, teachers and administrators of the Coventry, Connecticut, elementary schools. Special thanks go to Clarence Edmondson and Kathleen Osypuk of the George Hersey Robertson School, and to Mary M. Piazza and Annette Murphy of the Coventry Grammar School. We are grateful to Anne Fowler and two anonymous reviewers for their insightful comments and suggestions on an earlier draft of this paper.

REFERENCES

Baddeley, A. D., & Hitch, G. B. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and activation* (Vol. 4). New York: Academic Press.

Bohannon, J. N., Warren-Leubecker, A., & Hepler, N. (1984). Word order awareness and early reading. *Child Development, 55,* 1541-1548.

Bowey, J. A. (1986). Syntactic awareness in verbal performance from preschool to fifth grade. *Journal of Psycholinguistic Research, 15,* 285-308.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? *Applied Psycholinguistics, 2,* 201-212.

Chall, J. (1967). *Learning to read: The great debate.* New York: McGraw-Hill.

Chomsky, C. (1969). *The acquisition of syntax in children from 5 to 10.* Cambridge, MA: MIT Press.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 3,* 75-84.

Conrad, R. (1972). Speech and reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading.* Cambridge, MA: MIT Press.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusions, and memory span. *British Journal of Psychology, 55,* 75-84.

Crain, S., & Fodor, J. D. (forthcoming). Competence and performance in child language. In E. Dromi (Ed.), *Language and Cognition: A Developmental Perspective.* Norwood, NJ: Ablex.

Crain, S., & Shankweiler, D. (1988). Syntactic complexity and reading acquisition. In A. Davison & G. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered.* Hillsdale, NJ: Erlbaum.

Crain, S., Shankweiler, D., Macaruso, P., & Bar-Shalom, E. (in press). Working memory and sentence comprehension: Investigations of children with reading disorder. In G. Vallar & T. Shallice (Eds.), *Impairments of short-term memory.* Cambridge: Cambridge University Press.

Crowder, R. G. (1978). Audition and speech coding in short-term memory. In J. Requin (Ed.), *Attention and Performance, VII.* Hillsdale, NJ: Erlbaum.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466.

de Hirsch, K., Jansky, J., & Langford, W. (1966). *Predicting reading failure.* New York: Harper and Row.

Dunn, L., & Dunn L. (1981). Peabody Picture Vocabulary Test—Revised. Circle Pines, MI: American Guidance Service.

Fletcher, J. J., Satz, P., & Scholes, R. (1981). Developmental changes in the linguistic performance correlates of reading achievement. *Brain and Language, 13,* 78-90.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Forrest-Pressley, D. L., & Waller, T. G. (1984). *Cognition, metacognition and reading.* New York: Springer Verlag.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett.* Hillsdale, NJ: Erlbaum.

Fowler, A. (1988). Grammaticality judgments and reading skill in Grade 2. *Annals of Dyslexia* (Vol. 38, pp. 73-94).

Fry, M. A., Johnson, C. S., & Muehl, S. (1970). Oral language production in relation to reading achievement among select second graders. In D. J. Bakker & P. Satz (Eds.), *Specific reading disabilities: Advances in theory and method*. Rotterdam: Rotterdam University Press.

Goldsmith, S. (1980). *Psycholinguistic bases of reading disability: A study in sentence comprehension*. Unpublished doctoral dissertation, The City University of New York.

Goodluck, H., & Tavakolian, S. (1982). Competence and processing in children's grammar of relative clauses. *Cognition, 11*, 1-27.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj, II (Ed.), *Language development, Volume 1: Syntax and semantics*. Hillsdale, NJ: Erlbaum.

Hamburger, H., & Crain, S. (1984). Acquisition of cognitive compiling. *Cognition, 17*, 85-136.

Jorm, A. (1983). Specific reading retardation and working memory: A review. *British Journal of Psychology, 74*, 311-342.

Liberman, A. M. (1988). Reading is hard just because listening is easy. Paper presented at the Seventh International Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Liberman, I. Y. (1983). A language-oriented view of reading and its disabilities. In H. Myklebust (Ed.), *Progress in learning disabilities* (Vol. 5). New York: Grune & Stratton.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex, 18*, 367-375.

Liberman, I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. *Remedial and Special Education, 6*, 8-17.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. *Memory and Cognition, 8*, 329-335.

Mann, V. A. Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language, 11*, 627-643.

Mattis, S., French, J. H., & Rapin, I. (1975). Dyslexia in children and young adults: Three independent neuropsychological syndromes. *Developmental Medicine and Child Neurology, 17*, 150-163.

Menyuk, P., & Flood, J. (1981). Language development, reading/writing problems and remediation. *Bulletin of the Orton Society, 31*, 13-28.

Morice, R., & Slaghuis, W. (1985). Language performance and reading ability at 8 years of age. *Applied Psycholinguistics, 6*, 141-160.

Nelson, H. E., & Warrington, E. K. (1980). An investigation of memory function in dyslexic children. *British Journal of Psychology, 71*, 487-503.

Perfetti, C. A. (1985). *Reading Ability*. New York: Oxford University Press.

Richardson, E., & DiBenedetto, B. (1986). Decoding Skills Test. Parkton, MD: York Press.

Roit, M. L. (1980). *Syntactic development in normals and dyslexics: An empirical examination of two theoretical explanations of dyslexia*. Unpublished doctoral dissertation, Harvard University.

Rosenbaum, P. S. (1967). *The grammar of English predicate construction*. Cambridge, MA: MIT Press.

Rozin, P., & Gleitman, L. R. (1977). The structure and acquisition of reading II: The reading process and the acquisition of the alphabet principle. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading*. Hillsdale, NJ: Erlbaum.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorders: A modular approach. *Cognition, 24*, 139-168.

Shankweiler, D., Crain, S., Brady, S., & Macaruso, P. (in press). Identifying the causes of reading disability. In P. B. Gough (Ed.), *Reading acquisition*. Hillsdale, NJ: Erlbaum.

Shankweiler, D., & Liberman, I. Y. (1972). Misreading: A search for causes. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press.

Shankweiler, D., Smith, S. T., & Mann, V. A. (1984). Repetition and comprehension of spoken sentences by reading-disabled children. *Brain and Language, 23*, 241-257.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior, 13,* 272-281.

Smith, S. T., Mann, V. A., & Shankweiler, D. (1986). Spoken sentence comprehension by good and poor readers: A study with the Token Test. *Cortex, 22,* 627-632.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360-406.

Stein, C. L., Cairns, H. S., & Zurif, E. B. (1984). Sentence comprehension limitations related to syntactic deficits in reading-disabled children. *Applied Psycholinguistics, 5,* 305-322. .

Tavakolian, S. L. (1981). The conjoined-clause analysis of relative clauses. In S. L. Tavakolian (Ed.), *Language acquisition and linguistic theory.* Cambridge, MA: MIT Press.

Vellutino, F. R. (1979). Dyslexia: Theory and research. Cambridge, MA: MIT Press.

Vellutino, F. R., Pruzek, R. M., Steger, J. A., & Meshoulam, U. (1973). Immediate visual recall in poor and normal readers as a function of orthographic-linguistic familiarity. *Cortex, 9,* 370-386.

Vogel, S. A. (1975). Syntactic abilities in normal and dyslexic children. Baltimore, MD: University Park Press.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192-212.

Weinstein, R., & Rabinovitch, M. S. (1971). Sentence structure and retention in good and poor readers. *Journal of Educational Psychology, 62,* 25-30.

Wiig, E. H., & Roach, M. A. (1975). Immediate recall of semantically varied sentences by learning disabled readers. *Perceptual Motor Skills, 40,* 119-125.

FOOTNOTES

[*] *Applied Psycholinguistics,* in press.

[†] Also University of Connecticut, Storrs. Now at Neurolinguistics Laboratory, Institute of Health Professions, Massachusetts General Hospital, Boston.

[††] Also University of Connecticut, Storrs.

[1] The evidence that poor readers share in common difficulties in decoding words in print is overwhelming (see Chall, 1967; Perfetti, 1985; Stanovich, 1986). Evidence that the decoding problems are a reflection of wider deficiencies in processing the phonological components of language is summarized in several places (Liberman & Shankweiler, 1985; Rozin & Gleitman, 1977; Shankweiler & Liberman, 1972). Verbal working memory is another aspect of phonological processing that is deficient in poor readers (see Jorm, 1983; Liberman & Shankweiler, 1985; Wagner & Torgesen, 1987).

[2] It should be noted that the difficulties poor readers encounter in sentence comprehension would often be missed in everyday settings, but may come to light in tasks that present complex structures without the contextual support that is found in ordinary spoken discourse.

[3] In this study, a subject's score for each DST subtest consisted of the number of items read correctly prior to making ten errors in a row or completing the 60-item list. The maximum combined score was 120.

[4] Considerations governing the selection of an appropriate control group in investigations of the causes of reading disability are discussed in Shankweiler, Crain, Brady, & Macaruso (in press).

[5] Erroneous conjoined-clause interpretations of each type of relative clause sentence are described in Tavakolian (1981). For example, in SO sentence (7), a conjoined-clause misanalysis yields two possible interpretations: either the man kissed the lady and held an umbrella, or the lady kissed the man *and* held an umbrella. Tavakolian supposed that her preschool subjects may have applied preexisting phrase structure rules for conjoined clause sentences to relative clause sentences.

[6] Minimum distance errors are based on the minimum-distance principle which holds that a noun most proximal to the verb is interpreted as its missing subject (Chomsky, 1969; Rosenbaum, 1967) that were noted in the Mann et al. (1984) study, failed to distinguish between reader groups in the present study.

---

[7] For the SO sentences, in which the descriptive information is contained in the main clause, subjects in the present study acted out the *main* clause first 80.6% of the time (good readers: 74.7%; poor readers: 86.4%). This contrasts with the findings of a follow-up study by Crain, Shankweiler, Macaruso and Bar-Shalom (in press) in which the test sentences contained no descriptive information. In that study both the good and poor readers acted out the *relative clause* first 100% of the time for the SO sentences. Since, in the present study, the descriptive information appears in the relative clause for the SS, OS and OO sentences, the two response tendencies cannot be separated for these sentence types.

[8] It is important to note that the poor readers did *not* have more difficulty than the good readers with OS sentences, although in these sentences, too, the descriptive information followed mention of the overt action, as in the conjoined-clause sentences. The absence of reader group differences on OS sentences can be understood in light of the difference in pragmatic function of the two sentence types. Conceptually, the second conjunct of a conjoined-clause sentence is generally the (temporally) second event in an action sequence. In contrast, the relative clause, which appears second in an OS relative, has been found to denote the conceptually prior event in sentences which contain two action sequences (see Hamburger & Crain, 1982, for discussion).

[9] This explanation is developed within a framework that has long guided research on speech and reading at Haskins Laboratories (see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Related views of the language processing system have been advanced by Forster (1979) and Fodor (1983).

Phonological Deficiencies: Effective Predictors of Future Reading Problems

Virginia A. Mann[†] and Patricia Ditunno[†]

This chapter describes a recent longitudinal study which illustrates the role of phonological deficiencies in developmental reading problems. In that study, a population of children received a battery of phonological tests and a test of reading ability each year between kindergarten and first grade. The test battery was concentrated on three areas of phonological skill: 1) awareness of phonological structure, 2) retrieval and perception of phonological structure and 3) the use of phonetic representation in working memory, and it paired tests of these abilities with non-linguistic tests that controlled for extra-linguistic demands on attention, logical deduction, grapho-motor skill, etc.

The analysis of the results involved cross-longitudinal time lag correlations, partial correlations, regressions, and factor analysis as well as direct comparisons between good and poor readers. Three major conclusions are supported: 1) Phonological deficiencies play a causative role in early reading problems 2) In contrast, deficiencies in non-linguistic realms of cognition are not strongly associated with reading problems and 3) Poor readers tend to be "phonologically delayed"relative to their more successfully-reading peers, rather than "phonologically deviant."

In many different countries, certain spoken language skills are strongly and significantly related to early reading skill. Good and poor beginning readers have been distinguished by the status of these skills more reliably than by their general intelligence or other general cognitive abilities (for recent review, see, for example, Liberman, 1982; Mann, 1986a; Stanovich, Cunningham, & Freeman, 1984; Wagner & Torgesen, 1987). That children's ability to read should be related to their ability to use spoken language is anticipated by two major findings about the psychology of reading. First, there is the theoretical insight that writing systems are transcriptions of spoken language (see Liberman, Liberman, Mattingly, & Shankweiler, 1980). Second, there is the experimental observation that skilled reading is mediated by certain spoken language processing skills (see, for reviews: Crowder, 1982 or Perfetti, 1985, or Stanovich, 1982 a and b).

Haskins Laboratories SR-97/98

Status Report on Speech Research *1989*

In this chapter we will review some of the evidence that links reading ability to spoken language skills, and we will be describing in some detail a longitudinal study which shows how certain spoken language skills can actually presage the success with which a kindergarten child will learn to read in the first grade. The language skills which we will examine are phonological skills—those skills which operate on the basic sound elements of language and the regular patterns among them. These are of particular importance to beginning readers of alphabets (Liberman & Mann, 1980; Lundberg, Oloffson, & Wall, 1980; Mann, 1986a; Stanovich, Cunningham, & Cramer, 1984; Wagner &Torgessen, 1987), because alphabets are transcriptions of phonological elements (see Liberman et al., 1980). Above and beyond their role in decoding an alphabetic writing system, phonological processing skills are also important because they mediate all spoken language communication.

One phonological skill which relates to children's ability to read an alphabet is the ability to manipulate phonemes and syllables. An explicit (metalinguistic) awareness about phonemes is particularly critical to the child's realization of what the alphabet is "all about" because phonemes are more-or-less what the letters represent. This insight about the demands of an alphabetic writing system has led many investigators to consider beginning readers' awareness of phonemes as a factor in their reading success. They have examined children's reading behavior (Mann, 1984; Shankweiler & Liberman, 1972) and spelling errors (Liberman et al., 1986; Mann et al., 1987) as well as performance on various tests that directly measure the ability to count, delete or otherwise manipulate phonemes and syllables (another aspect of phonological structure). All of these diverse approaches have shown that insufficient awareness about phonemes is a common trait of poor beginning readers whereas superior awareness about phonological structures associates with superior reading ability (for reviews, see: Liberman, 1983; Mann, 1986a; Mann & Liberman, 1984; Mattingly, 1984; Stanovich, Cunningham, & Cramer, 1984; Treiman & Baron, 1981, or Perfetti, 1985).

Other, more fundamental, phonological skills are also linked to early reading, and they involve the ability to process spoken language. Several types of phonological processing skills are now known to distinguish poor readers from good readers. For example, reading problems often associate with a problem in retrieving and perceiving the phonological structure of words. With regard to the retrieval of phonological structures, good readers make fewer errors than poor readers when naming letters or pictures—tasks which require them to retrieve spoken words from their mental lexicon (Katz, 1986; Wolf and Goodglass, 1986). With regard to perception, good readers' perception of speech in noise tends to be superior to that of poor readers, although the two groups do not differ in audiometry and perform at more or less the same level when the test items are nonlinguistic sounds such as a cat meowing or a door slamming (Brady et al., 1983). Thus the problems of poor readers appear to be confined to the linguistic domain.

Another area of phonological processing which is linked to reading skill has been identified by studies of short-term (working) memory (see, for example: Brady, 1986; Jorm, 1979; Mann et al., 1980; Mann & Liberman, 1984). It has been shown that good readers surpass poor readers in the ability to recall sequences of letters, whether written or spoken (Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). Good readers also surpass poor readers in the ability to repeat strings of spoken words (Brady et al., 1983; Mann et al., 1980) and even in the ability to repeat meaningful sentences (Mann et al., 1980). However, studies which have examined the recall of non-linguistic material such as unfamiliar faces, nonsense drawings or visual-spatial sequences on the Corsi block test reveal no differences between good and poor readers (see Liberman et al., 1982 and Katz, Shankweiler, & Liberman, 1981, for example). It is only in those cases where the to-be-remembered material is either spoken utterances or nameable visual stimuli that we tend to find significant differences between good and poor readers.

Further studies of the pattern of children's responses when they attempt to hold linguistic material in memory, and their susceptibility to a manipulation of phonological structure (i.e., rhyme) offer an explanation of the poor readers' difficulty with holding phonetic information in working memory. These children may fail to recall linguistic material as well as good readers do because they encounter some difficulty with using a speech code (phonetic representation) in working memory (see, for a review Brady, 1986; Mann, 1986a). The impact of this working memory problem is quite far reaching; it leads to problems in such "higher" levels of language use as sentence comprehension (Crane & Shankweiler, 1986; Mann, Smith, & Shankweiler, 1984).

In summary, then, poor readers in the early elementary grades can be distinguished from good readers by phonological deficiencies in at least three areas 1) awareness of phonological structure 2) retrieval and perception of phonological structures and 3) use of phonetic representation in working memory. In further studies of these deficiencies and their relation to reading skill, some investigators have recognized the need for longitudinal research which asks whether phonological deficiencies presage reading problems. Longitudinal research is an attractive approach because it both validates and enhances research into the basis of reading disability. It is the only effective means of determining whether individual differences in phonological skills actually cause differences in reading ability, as opposed to being their consequence. It also addresses two very practical needs that teachers and school systems must face—the need for preschool screening and the need for more effective remediation and enrichment.

To date, of the many studies that have concerned early reading problems, relatively few have used a longitudinal design to probe the determinants of those problems. Those few studies, however, have been quite promising. They have revealed that children who become poor readers in the first grade tend, as kindergartners, to be less aware of phonological structure (Blachman, 1983; Bradley & Bryant, 1985; Mann & Liberman, 1984; Stanovich et al., 1984), less able to perform naming tasks (Mann, 1984, Wolf & Goodglass, 1986) and less able to repeat strings of words (Mann, 1984; Mann & Liberman, 1984). Children who become good readers excel in each of these three areas and average readers fall somewhere between.

What has been lacking is a longitudinal study which systematically examines the relation between reading ability and the three areas of phonological awareness, phonological retrieval and perception. No study has examined all three areas. Also, many studies have lacked effective controls for kindergarten reading ability and time of testing, and many have also failed to compare non-linguistic tests to the linguistic tests which seem to predict reading ability. Such controls are essential to the conclusion that the phonological demands of a given test are the basis for its predictive power; they are the best means of excluding demands on attention, logical deduction, grapho-motor skill etc. as confounding factors.

There has also been little attempt to discern the basis of the poor readers' problems. For example, their language difficulties could reflect some type of maturational delay (as discussed by Mann, Cowin, & Schoenheimer, in press; Mann & Liberman, 1984) as opposed to a real deviance in development. Delay or deviance could be specific to language, or could extend to non-linguistic areas of cognition as well. The possibility that poor readers are 'phonologically delayed' with respect to other children has received some support from a recent study conducted in our lab (Mann et al., in press). Further consideration of this possibility will require comparisons of the differences between good and poor readers and those that exist between older vs. younger children.

Longitudinal research makes such comparisons readily available. These considerations led us to conduct a study which addressed three questions: 1) Do phonological deficiencies actually cause reading problems as opposed to being their consequence? 2) When the extraneous demands of individual tests are taken into account, do phonological deficiencies remain the best effective predictors of future reading problems? and 3) Should poor readers be characterized as "phonologically delayed," or are they "phonologically deviant" relative to normal and superior readers? We now turn to a detailed report of the methodology and the analysis of the results.

Method

Subjects. All subjects were recruited from three public elementary schools in the northwestern suburbs of Philadelphia. They participated with the permission of their parents and at the convenience of their teachers.

The design called for two years of longitudinal testing beginning with the kindergarten year. At the onset of the study, 106 kindergarten children participated, and they were divided into two groups. One group of 51 children was tested in the fall of each year (between the second and fourth months of the school year), and the other group of 55 children was tested in the spring (between the ninth and tenth months).

The following year, 70 of these children were available for first grade testing, 31 in the fall group and 39 in the spring group. Children tested in the fall of kindergarten were seen in the fall of first grade; those tested in the spring of kindergarten were tested in the spring of first grade.

After the study had been completed, 50 of the children were available for a reading test which was administered at the end of the second grade. The results of this test appear in the tables, although we will focus our analysis on the prediction of first-grade reading ability.

Materials

Each year between kindergarten and first-grade, we administered a battery of tests to determine whether certain phonological skills predicted future reading ability. The battery included five different phonological tests and five non-linguistic control tests that made similar demands on attention, logic, motor skills etc. without demanding linguistic skills, per se. Two linguistic tests examined children's *awareness of phonological structure*: a test of syllable awareness (as employed in Mann & Liberman, 1984), and an invented spelling test (as discussed in Mann et al., 1987). The nonlinguistic controls for these tests included a test of angle awareness (from Mann, 1987) and the Goodenough Draw-a-Man test (Harris, 1963). Two linguistic tests examined *retrieval and perception of the phonological structure of words*: a speeded letter naming test (from Mann, 1984) and a test of word perception in noise (from Brady et al., 1983). The nonlinguistic controls for these included a test of environmental sound perception in noise (also from Brady et al., 1983) and the Denckla test for speeded fine motor coordination (Denckla, 1973) which was only given in Phase II. The final linguistic test was a word string recall test (the nonrhyming word strings from Mann, 1984), which examined the ability to make *use of phonetic representation in working memory*. The control for this test was a test of memory for visual-spatial sequences on the Corsi blocks (as discussed in Mann & Liberman, 1984).

In addition to the experimental test battery, we also administered a reading test each year of the study (i.e., the Word Identification and Word Attack subtests of Form A of the Woodcock Reading Mastery Tests; Woodcock, 1973). A measure of IQ (the vocabulary and block design subtests of the WPPSI) was also taken during kindergarten testing. A detailed description of each part of our battery follows, with each non-linguistic control tests described following the phonological test that it concerns.

1. Awareness of Phonological Structure

1.1 Syllable Counting

This test was a shortened version of the syllable counting test developed by Liberman et al. (1974) and employed in Mann and Liberman (1984). The child was asked to deduce the rules of a "counting" game which involved using markers to count the number of units (syllables) in spoken words. The instructor presented four different series of three training items like "but, butter, butterfly" which systematically illustrated that one syllable words receive one marker, two syllable words receive two markers, etc., without referring to the term "syllable." Feedback was given until the child could correctly "count" each item in each of the training series, 21 test items then followed without response feedback (i.e., the first 21 of the 42 items from the Liberman et al. test). Scoring was as in Mann and Liberman (1984): the total number of correct items (ASC, Max.=21) and a pass/fail score (ASPF, Max.=1) in which pass denoted at least five correct responses in a row.

1.2 Angle Counting

This test was analogous to those in the syllable counting test described above. The only difference was that the items were drawings which contained between one and three thirty degree angles embedded within their contours; the child's task was to count the number of thirty degree angles in each drawing. The four training trials presented drawings which varied progressively along a dimension of complexity from explicit angles to angles embedded in figures. The 21 test items presented drawings of familiar scenes and objects. Scoring was the same as in the syllable awareness test, the total number of correct items (AAC, Max. = 21) and a pass/fail score (AAPF, Max. = 1).

1.3 Invented Spelling

This test was recently described in Mann, Tobin and Wilson (1987). In it, children were presented with fourteen words to which young children often give "invented" spellings that reflect a growing appreciation of phonological structure (as discussed in Read, 1986). The children were asked to try to "write" each word as best they can (or to draw a picture if they cannot write it), and the instructor encouraged them to try to write as many sounds as they hear. The scoring of each word involved a four point scale: 4 points were given to words that were correctly spelled, but up to 3 points were given to unorthodox spellings that captured the phonological structure of the word (as in "PPL" for "people"). The score was the sum of points across individual words (ISC, Max. = 56).

1.4 Draw a Man

This test, taken from Harris (1963) required children to draw a human figure. Scoring considered the accuracy of the drawing they produce (DAM, Max. = 73). described in Harris (1963).

2. Retrieval and Perception of Phonological Structures

2.1 Speeded Letter Naming

This test, taken from Mann (1984), presented children with a five-by-five array of capital letters arranged in a random sequence. The child was required to name the letters, row by row as quickly and as accurately as possible. Two scores were computed, the elapsed time (LNT) and the number of errors (LNE, Max. = 25).

2.2 Denckla fine-motor coordination

As described in Denckla (1973), this test required children to make successive movements of the thumb and fingers of the dominant hand. After the child had practiced the required movement, the instructor determined the time required to make 5 complete sets of movements (1 set = index finger to thumb, middle finger to thumb, etc.) The score was the elapsed time (DFTT). This test was introduced into the study after the kindergarten phase of testing had been completed.

2.3 Perception of Spoken Words in Noise

This test, adapted from the materials of Brady et al. (1983), consisted of 24 high frequency words, each masked by signal correlated noise. These were divided into two lists of 12 words each, one for use in kindergarten and one for use in first grade. Each list was preceded by two practice items; children's task was to identify each word. Three points were given for each correct item and one point if the vowel alone was correct. The score was the sum of points across words (PWN, Max. = 36).

2.4 Perception of Environmental Sounds in Noise

This test, also adapted from Brady et al. (1983), included 24 environmental sounds that had been masked in broad-band white noise. Twelve of the stimuli were used for kindergarten testing and twelve were used in first grade. Practice items preceded the test items, and children were asked to identify each sound (i.e., a plane, an orchestra, a cat, etc.) Each correct response merited 3 points, and a scoring system (as in Brady et al., 1983) gave 0, 1 or 2 points according to how closely an incorrect response matched the correct one. The score was the sum of points across stimuli (PSN, Max. = 36).

3. Use of Phonetic Representation in Working Memory

3.1 Word String Recall

The materials for this test were taken from Mann (1984) and consisted of seven word strings, each four words long. The words were drawn from the A and AA sets of the Lorge-Thorndike count and each occurred only once. For the purpose of testing, the strings were prerecorded by a female native speaker of English who kept intonation neutral and spoken one per second. During testing, the child was told that he/she would hear four words, and that he/she should remember them and repeat them in order as soon as the fourth word was heard. Responses were scored in terms of the number of correct responses, summed across the seven word strings. Following Mann et al. (1980), two scores were computed, one was independent of order (WSF, Max. = 28) and the other was order-strict (WSS, Max. = 28).

3.2 Corsi Block Recall

This test, as described in Mann and Liberman (1984), was administered with a set of nine black cubes scattered about a flat black base. Each block was identified by a number which only the experimenter could see; during testing, predetermined sequences of blocks were "tapped" at the rate of one per second; the child watched and attempted to reproduce each sequence. Testing began with four sequences of two blocks, followed by four of three blocks and four of four blocks. Two scores were computed, an order independent score (CBF, Max. = 36) and an order-strict score (CBS, Max. = 36).

Procedure

Each year, testing began with group administration of the Invented Spelling Test and the Draw-A-Man test, then proceeded to individual administration of the remaining tests which were given in a fixed, counterbalanced order over the course of two experimental sessions. In the first session, the WPPSI subtests were administered (in kindergarten only), followed by one of the "perception-in-noise" tests, one of the "counting" tests and the Denckla test I (in first grade only). In the second session, the letter naming test, the reading tests, the word string recall test, the remaining "perception in noise" test, the Corsi test and the remaining "counting" test were administered.

Results

The results are summarized in Tables 1-5 and in Figures 1-3. Examination of Table 1 will indicate the appreciable gains in reading ability that children achieved over the course of the study and some analogous gains in performance on the test battery. The primary question to be asked is whether performance on the phonological tests predicts future reading ability. To this end, we first report an analysis of the data which has involved: 1) Time-lag and Partial Correlations, 2) Multiple Regressions and 3) Factor Analysis. A second question concerns the characterization of poor readers' problems with phonological tests, and the possibility that they exhibit some type of maturational delay. In seeking to answer this question, we will make comparisons between good and poor readers, relating their scores to the mean score that was obtained in each phase of the study (i.e., fall of kindergarten, spring of kindergarten, fall of first grade, spring of first grade).

TABLE 1. Mean Scores (and standard deviations) on Reading Tests and on the Experimental Battery at Each Time of Testing.

		Kindergarten		First Grade	
		Fall	Spring	Fall	Spring
Reading Ability:					
Word ID		1.55 (5.05)	8.11 (20.51)	17.86 (22.10)	71.89 (21.96)
Word Attack		0.00 (0.00)	0.00 (0.00)	6.60 (6.69)	24.96 (15.30)
Phonological Awareness:					
Counting tests					
Linguistic:	ASC	7.28 (2.77)	8.11 (4.31)	10.27 (4.13)	12.49 (5.22)
	ASPF	0.10 (0.30)	0.15 (0.36)	0.40 (0.50)	0.55 (0.50)
Control:	AAC	7.90 (2.37)	7.56 (2.45)	7.80 (2.22)	7.53 (2.49)
	AAPF	0.10 (0.30)	0.09 (0.29)	0.11 (0.32)	0.22 (0.42)
Phonological Awareness:					
Invented Spelling					
Linguistic:	ISC	12.36 (7.67)	20.60 (12.04)	36.52 (9.78)	51.54 (4.72)
Control:	DAM	15.61 (5.59)	20.44 (6.23)	16.20 (3.87)	20.58 (5.65)
Phonological retrieval:					
	LNT	61.16 (42.65)	41.67 (28.49)	32.76 (22.67)	16.82 (4.20)
	LNE	4.67 (4.67)	1.42 (1.42)	0.93 (1.20)	0.09 (0.29)
Phonological Perception:					
Linguistic:	PWN	16.39 (1.74)	15.86 (2.35)	16.13 (1.68)	19.84 (2.25)
Control:	PSN	9.80 (3.72)	6.36 (2.96)	10.56 (3.02)	10.33 (2.62)
Phonological Working Memory:					
Linguistic:	WSF	20.22 (4.50)	20.56 (4.86)	23.78 (2.98)	23.56 (3.44)
	WSS	13.98 (7.71)	15.76 (8.29)	19.58 (5.83)	20.53 (0.83)
Control:	CBF	32.90 (2.01)	33.87 (2.13)	34.87 (1.27)	35.78 (1.12)
	CBS	27.33 (4.53)	29.54 (4.80)	31.71 (3.21)	32.46 (2.64)
WPPSI subtests:					
Vocabulary		16.98 (3.45)	18.84 (4.10)		
Block design		12.22 (3.23)	14.93 (2.49)		
IQ		104.65 (8.34)	103.91 (6.40)		

Cross-lag and Partial Correlations

There are several ways to determine whether performance on a given test predicts future reading ability. One is to use longitudinal *cross-lag correlations*, a form of analysis which asks whether kindergarten scores on the experimental battery bear a stronger relationship to first grade reading ability (forward) than first grade scores on that battery skills bear to kindergarten reading ability (backward). Do phonological skills predict reading ability to a higher degree than reading ability predicts phonological skills? Table 2 indicates that, for seven of the eight linguistic tasks the forward-directed "predictive" relation is indeed a higher value than the reverse-directed "control" relation. These tests include seven out of the eight measures of phonological skill: the two measures of syllable counting performance (ASC and ASP), the invented spelling test of phonological awareness (ISC), the two measures of letter naming ability (LNE and LNT which is only greater in the case of the fall testing), the two measures of word-string recall (WSOF and WSOF which is only greater in the case of spring testing). In contrast, only one of the seven nonlinguistic control tests is more predictive of reading ability than vice versa (i.e., the order-strict Corsi block test, WSOS) and this is apparent only in the spring sample. For one test (the Draw-a-Man test, DAM) the control relationship was actually stronger than the predictive one, and in all other cases the relations were either equivalent or non-significant.

A second means of evaluating the predictive power of a test is to consider partial correlations. Performance on many of the tests within the experimental battery was related to performance on the Woodcock subtests (measured in terms of the combined number of correct responses), as can be seen in Tables 3 and 4. However, kindergarten reading ability is also related to first grade reading ability and to performance on many of the battery tests. For this reason, it is possible that any correlations between the kindergarten scores on the test battery and first grade reading ability are by-products of kindergarten reading ability; phonological skills could be consequences of reading rather than causes. To control for this possibility we can set a criterion, following Perfetti (1985) and Wagner and Torgesen (1987) that any test which truly predicts reading ability should remain correlated with first grade reading ability when the contribution of kindergarten reading ability is removed by the technique of partial correlations. Many of the phonological tests in our battery pass this control, which we take as evidence that they are true predictors of future reading ability.

Specifically, the tests which remain significantly correlated at $p <.05$ include the fall testing of syllable awareness (both ASC r=.41, $t(28)$=2.38 and ASPF, r=.37, $t(28)$=2.11), invented spelling (r=.59, $t(28)$=3.86) and naming speed (both speed r= -.36, $t(28)$=2.12 and errors r=-.40, $t(28)$=2.12). In the fall testing, the ordered word string recall fell just short of significance (r=.29 t(28)=1.6). For spring testing, the predictive tests included the tests of syllable awareness (both ASP, r=.51, $t(36)$=3.28 a,d ASPF, r=.51, $t(36)$=2.62), the invented spelling test (r=.34, $t(36)$=2.17), letter naming speed (r=-.41, $t(36)$=2.84), and the word string recall (ordered score, r=.40 $t(36)$=2.88). The WPPSI vocabulary test was also predictive in both fall (r=.25, $t(28)$=2.76) and spring $t(36)$=2.18, r= .33), although the block design test was not. Finally, the Corsi block ordered score was significantly related to future reading ability in the spring testing, (r=.35, $t(36)$=2.24), but not in the fall, consistent with the time-lag analysis.

TABLE 2. Predictive vs. Control Correlations: Relation between the Experimental Test Battery and Reading Ability.

		Fall Testing		Spring Testing	
		Predictive	Control	Predictive	Control
Phonological Awareness: Counting tests					
Linguistic:	ASC	.50	.06	.55	.31
	ASPF	.47	.25	.59	.28
Control:	AAC	.27	.07	.28	.13
	AAPF	.22	.07	.16	.04
Phonological Awareness: Writing & Drawing					
Linguistic:	ISC	.71	.49	.41	.32
Control:	DAM	.27	.42	.23	.46
Naming Speed					
Linguistic:	LNT	-.39	-.10	-.25	-.38
	LNE	-.41	.10	-.47	-.08
Control:	DFTT	NA	.06	NA	.05
Sound Perception					
Linguistic:	PWN	.02	.04	.11	.27
Control:	PSN	-.11	.03	.04	.04
Working Memory					
Linguistic:	WSF	.36	.02	.52	.23
	WSS	.16	.09	.54	.30
Control:	CBF	.09	.08	.09	.12
	CBS	.26	.10	.35	.18
Reading ability		.56	.56	.51	.51
N		31	31	39	39
Significance levels					
$p < .05$		.31	.31	.27	.27
$p < .01$		.42	.42	.37	.37

One phonological test is noteworthy for its lack of significance according to either correlational analysis: the test of speech perception in noise. Apparently, that test was not a very effective kindergarten predictor of future reading ability. However, the data that appear in Table 4 suggest that this test becomes predictive when it is given in first grade. First grade scores on speech perception in noise are positively correlated with future reading ability, and a negative correlation is obtained between first grade perception of environmental sounds in noise and future reading ability. Both of these correlations are consistent with the findings of Brady et al. (1983), who reported significant differences between good and poor readers in a third grade sample. The fact that this result did not obtain among the kindergartners we tested might suggest that the predictive power of a given test may depend on the age at which it is administered and the age at which reading ability is assessed.

TABLE 3. Correlations between the Test Battery and Reading Ability: Kindergarten Administration.

		Fall Testing			Spring Testing		
Reading Ability in:		Kdgn	1st	2nd	Kdgn	1st	2nd
Phonological Awareness: Counting tests							
Linguistic:	ASC	.33	.50	.07	.31	.55	.28
	ASPF	.33	.47	.23	.62	.59	.31
Control:	AAC	.14	.27	.31	.09	.28	.15
	AAPF	.11	.32	.35	.08	.16	.09
Phonological Awareness: Writing & Drawing							
Linguistic:	ISC	.54	.71	.44	.26	.41	.31
Control:	DAM	.26	.27	.17	.21	.23	.29
Naming Speed							
Linguistic:	LNT	-.21	-.39	-.43	-.25	-.25	-.33
	LNE	-.22	-.41	-.68	-.19	-.47	-.48
Control:	DFTT	NA	NA	NA	NA	NA	NA
Sound Perception							
Linguistic:	PWN	.10	.02	.13	.28	.11	.16
Control:	PSN	-.04	-.11	.22	.27	.04	.23
Working Memory							
Linguistic:	WSF	.24	.36	.06	.37	.52	.26
WSS		.09	.16	.18	.37	.54	.33
Control:	CBF	.03	.09	.20	.08	.09	.16
CBS		.07	.26	.17	.12	.35	.20
WPPSI							
Vocabulary		.27	.44	.52	.34	.33	.35
Block Design		.43	.51	.40	.19	.39	.34
IQ		.37	.52	.57	.23	.32	.30
Age		.30	.26	.08	.17	.13	.07
Gender		-.14	-.05	-.15	.03	.12	-.01
Reading Ability		NA	.56	.30	NA	.51	.36
N		51	31	18	55	39	32
Significance level							
$p < .05$		.23	.31	.40	.23	.27	.31
$p < .05$		.32	.42	.54	.32	.45	.42

TABLE 4. Correlations between the Test Battery and Reading Ability: First Grade Administration.

		Fall Testing			Spring Testing		
Reading Ability in:		Kdgn	1st	2nd	Kdgn	1st	2nd
Phonological Awareness: Counting tests							
Linguistic:	ASC	.06	.04	.13	.31	.45	.39
	ASPF	.25	.20	.34	.28	.41	.20
Control:	AAC	.07	.08	.22	.13	.06	.08
	AAPF	.07	.09	.17	.04	.02	.03
Phonological Awareness: Writing & Drawing							
Linguistic:	ISC	.49	.45	.85	.32	.74	.75
Control:	DAM	.42	.02	.39	.46	.16	.30
Naming Speed							
Linguistic:	LNT	-.10	-.38	-.82	-.38	-.50	-.67
	LNE	.10	-.42	.72	-.08	-.10	.04
Control:	DFTT	-.06	-.14	-.27	.05	-.12	.04
Sound Perception							
Linguistic:	PWN	.04	.47	.47	.27	.36	.21
Control:	PSN	.03	-.01	-.33	.04	-.23	-.34
Working Memory							
Linguistic:	WSF	.02	.03	-.07	.23	.40	.25
	WSS	.09	.12	.04	.30	.45	.23
Control:	CBF	.08	.06	.06	.12	.16	.05
	CBS	.10	.21	.17	.18	.38	.20
Reading Ability		.56	NA	.52	.51	NA	.83
N		31	31	18	39	39	32
Significance level							
$p < .05$		.31	.31	.40	.27	.27	.31
$p < .01$		.42	.42	.54	.37	.37	.42

Multiple Regression

Multiple regressions represent another means of determining whether kindergarten performance on the test battery predicted first grade reading ability. For the Fall and Spring testing, we have computed a multiple regression on reading ability, using the tests that were identified by the partial correlation analyses. These indicate that performance on the phonological tests accounts for approximately 60% of the variance in reading ability. Among children tested in the fall, the multiple correlation coefficient is .78, among those tested in the Spring the multiple correlation coefficient is .76.

More specifically, for Fall testing, the invented spelling test score accounted for 51% of the variance in reading ability, syllable awareness (pass fail score) accounted for an additional 7%, letter naming speed accounted for an additional 1% and word string memory for an additional 2%. When these are accounted for, performance on the

WPPSI vocabulary contributes less than 1% and kindergarten reading ability contributes an additional 15%. In the Spring group, performance on the angle counting test (pass fail score) account for 35% of the variance and performance on the invented spelling test accounts for an additional 4%. Memory for the word strings accounts for an additional 13%, errors in letter naming for an additional 4% and speed of letter naming for an additional 2%. When the contribution of the phonological tests is accounted for, performance on the Corsi block test contributes an additional 11%, WPPSI vocabulary contributes 2%, and kindergarten reading ability contributes 2%.

Factor Analyses

Factor analysis can offer yet another way of examining the relation between performance on the experimental battery and reading ability. For each time of testing, we have computed a varimax rotated factor analysis that includes kindergarten performance on the experimental battery, WPPSI performance, age, gender and both kindergarten and first grade reading ability. In the fall sample, the analysis converged on 9 factors after 13 iterations, in the spring, on 9 after 21. Table 5 lists the factors which were relevant to reading scores at a level of > .20, and also gives the other tests which loaded on each factor (at a level > .20).

TABLE 5. Factor Analysis: Kindergarten Performance and Reading Ability.

FALL			
I		II	
Word ID	.4873	Word ID	.2232
Word Attk	.4971	Word Attk	.1490
Kdgn Word ID	.2337	Kdgn Word ID	.2698
LNT	-.8203	MWF	.8983
LNE	-.8140	MWS	.8368
ISC	.6958	gender	-.6907
IQ	.6891		
WPPSIV	.6878		
WPPSIP	.6205		
Age	.2350		
DAM	.3429		
MNOS	-.2029		

SPRING							
I		II		III		IV	
Word ID	.5250	Word ID	.4782	Word ID	.3482	Word ID	.3001
Word Attk	.5536	Word Attk	.2473	Word Attk	.1381	Word Attk	.2432
Kdgn Word ID	.3730	Kdgn Word ID	.1391	Kdgn Word ID	.3303	Kdgn Word ID	.2012
ASC	.8521	MWOS	.7204	MBF	.3459	Age	.6263
ASPF	.8288	LNT	-.7128	MBS	.7451	ASPF	-.4350
ISC	.7105	MWF	.6274	ASC	-.4312		
LNE	-.3907	AAPF	.3303				
PSN	-.3810						

In general, it can be seen that the factors which related to reading ability also related to performance on the phonological tests, and this relationship was more systematic than relationships involving performance on the non-linguistic control tests. For the Fall testing, only two factors are related to reading ability: The primary factor for reading ability is also the primary factor for the tests of phonological awareness and letter naming ability and also for the WPPSI vocabulary test, the second factor relates to word string memory. For the spring testing, four factors are related to reading ability: The primary factor is again the primary factor for performance on the tests of phonological awareness, a second factor relates to letter naming ability and word string memory, a third relates to Corsi block performance and a fourth is related to age.

Comparisons of Good and Poor Readers

Aside from evaluating the possibility that reading disabilities are caused by phonological deficiencies, our analysis was designed to consider the possibility that poor readers are "delayed" in the acquisition of phonological skills. To this end, we asked the children's teachers to identify the "good" readers who were reading more than one grade level above expectation, and the "poor" readers who were reading more than one grade below expectation. This yielded 7 poor readers and 9 good readers in the fall sample, and 6 poor readers and 6 good readers in the spring sample. Although the samples are small in size, we may still pose certain questions about the performance of the poor readers in relation to that of good readers and in relation to slightly younger children. If the poor readers resemble slightly younger children in their performance on the experimental tests, it would be consistent with a "maturational delay" account of reading ability (as discussed in Mann, 1986). To the extent that good readers resemble slightly older children, it would support a "maturational acceleration" account of children who are "gifted" readers. If, however, the behavior of poor readers is considerably worse than all other children and follows a different course of development, we may consider the possibility of deviant development.

In Figures 1-3 we have plotted the performance of the good and poor readers in the Fall and Spring of first grade against a reference curve which is the mean performance of children at each time of testing (i.e., Fall of Kindergarten, Spring of Kindergarten, Fall of First Grade and Spring of First Grade). Figure 1a presents the predictions of a "maturational delay" account of poor readers performance (and a maturational acceleration account of good readers' superiority). That figure also plots the children's performance on the two reading tests and it will be noted that the reading ability of the poor readers is quite consistent with a maturational delay account. The plots for the tests within the experimental battery appear in Figures 2 and 3, where Figure 2 presents the results obtained with the linguistic tests and Figure 3, the non-linguistic ones. In general, the results obtained with the linguistic tests are in keeping with the predictions of the maturational account: The good readers tend to perform at a higher level than average whereas the poorer readers tend to perform at a lower level. Also consistent is the observation that, within each reading group, there tends to be improvement with age and the extent of the effect of age is much the same as that seen in the reference curve. In contrast, the results obtained with the non-linguistic tests are much less systematic: The effects of age are less clear cut and only the Draw-A-Man and Corsi block test gives the result predicted by the "maturational lag" account. Hence there are mixed results as to whether the delay is specific to language development, and more research is needed to address this point.

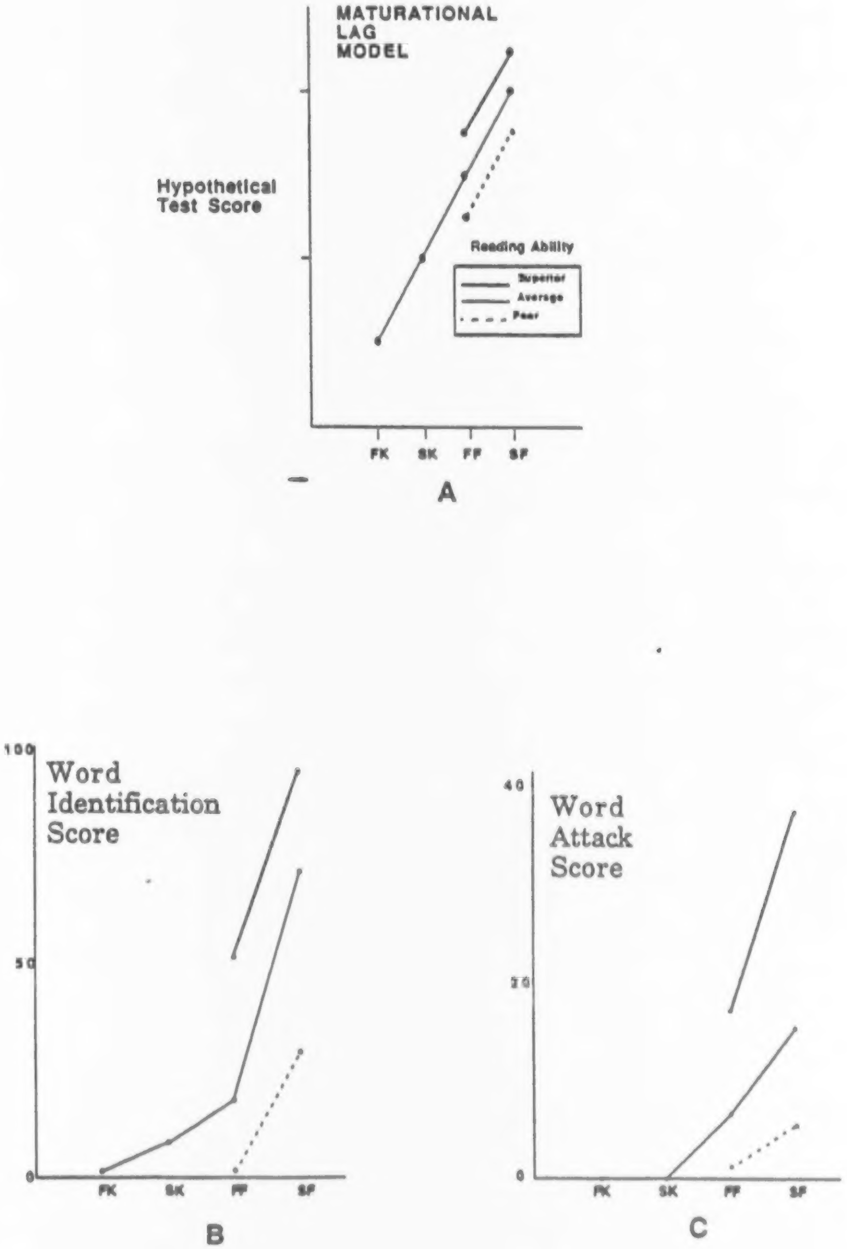


Figure 1. Comparisons of good and poor readers at two times of testing, in relation to the performance of normal children. A) Predictions of a "maturational delay" account of reading disability. B) Children's ability to identify words. C) Children's ability to read phonologically plausible nonwords.

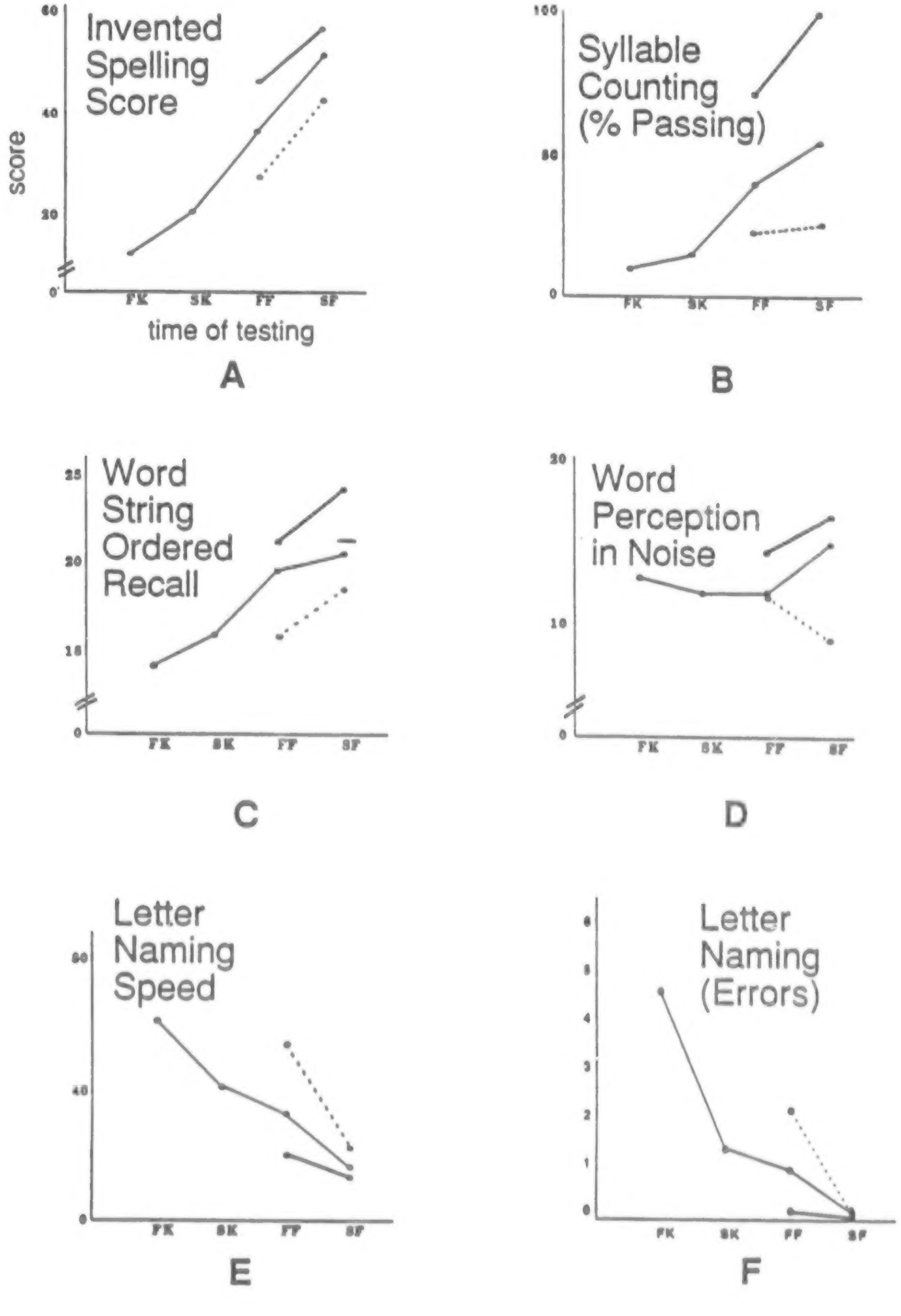


Figure 2. Comparisons of the language skills of good and poor readers at two times of testing, in relation to the performance of normal children. A) The invented spelling test of phonological awareness. B) The syllable counting test of phonological awareness. C) Use of phonetic representation in working memory. D) Perception of words in noise. E) Speed of lexical retrieval. F) Errors in lexical retrieval.

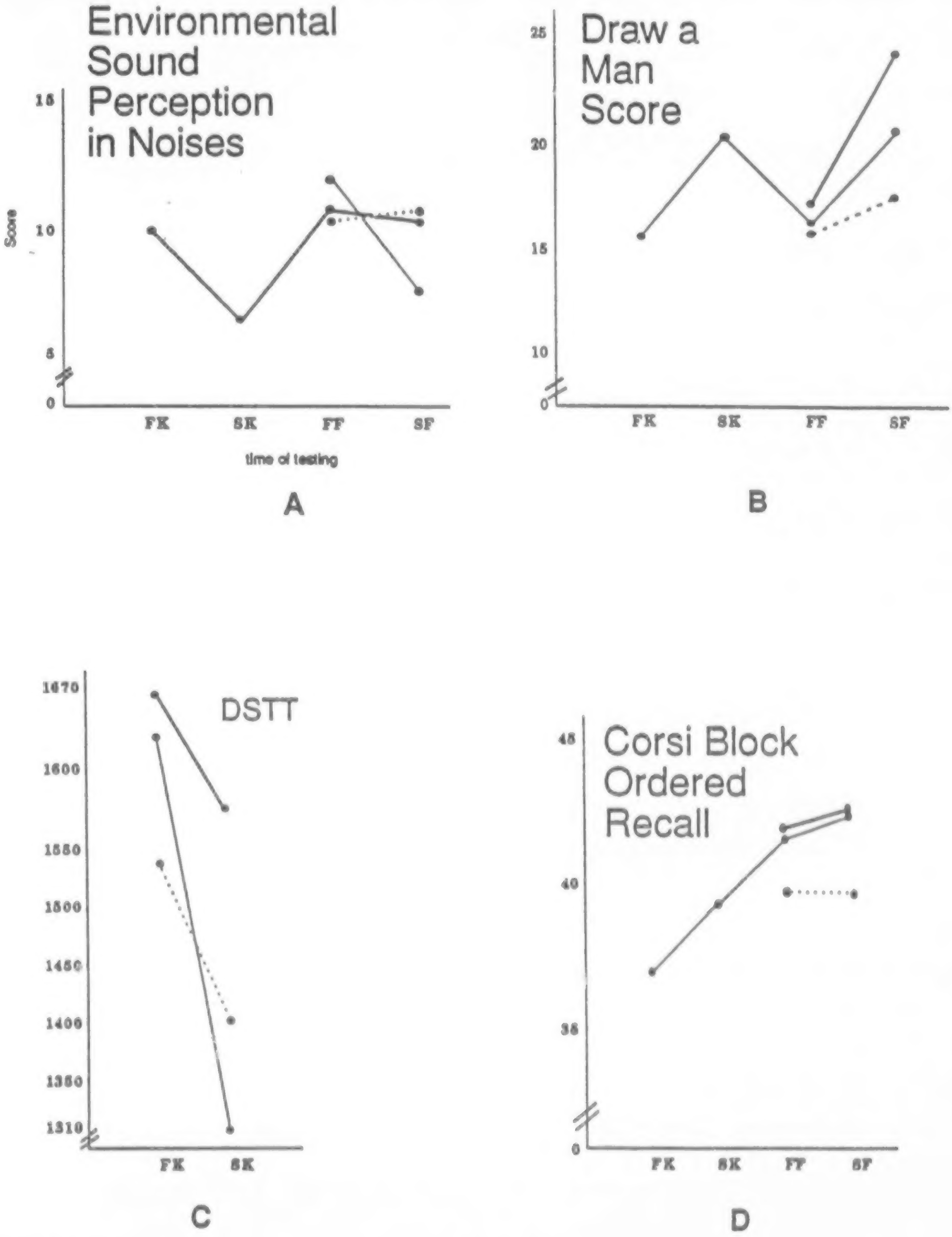


Figure 3. Comparisons of the non-linguistic skills of good and poor readers at two times of testing, in relation to the performance of normal children. A) Environmental Sound Perception. B) Draw-A-Man Ability. C) Speed of Repetitive Finger Tapping. D) Non-linguistic short-term memory.

Summary and Concluding Remarks

The findings of this study can be summarized in terms of three major points. The first and most important of these is that future reading problems can be predicted by deficiencies in each of the three areas of 1) awareness of phonological structure, 2) retrieval and perception of phonological structure and 3) use of phonetic representation in working memory. The cross-lag correlations between reading ability and performance on the experimental battery reveal that such phonological deficiencies predict future reading problems more successfully than deficient reading ability predicts future phonological problems. Even in the case of letter naming, a skill where one might expect reading experience to have a strong effect, the ability to name letters rapidly is a better predictor of reading ability than vice versa. Likewise, the partial correlations indicate that, even when kindergarten reading ability is taken into account, phonological skills remain effective predictors of first-grade reading ability. Skills involved with memory, retrieval and awareness of phonological structure are more than concomitants of reading ability: they actually presage it. Indeed, the results of the multiple correlation analysis reveal that these skills may account for up to 60% of the variance in children's future reading ability.

To some extent, the three areas of memory, retrieval and awareness appear inter-related, yet there are some indications that they involve separable factors. Those indications follow from a consideration of the factor analysis, which revealed that phonological skills tended to load onto two different factors, each of which also loads onto reading ability. One factor includes skills concerned with the awareness and retrieval of phonological structure which appear to be related; the other concerns temporary memory processes. Retrieval processes load on factors associated with either awareness or memory, depending on whether the fall or spring group is at issue. Thus it would seem that phoneme awareness skills and verbal short-term memory skills make a somewhat independent contribution to reading ability and this is consistent with some previous indications (Mann & Liberman, 1984) about the relation between these skills and reading ability. These indications, together with the results of the multiple correlation, suggest that an effective predictive battery should include multiple measures of phonological skill.

A second major point concerns the fact that poor performance on tests of phonological skills is a considerably more consistent and effective predictor of future reading problems than is performance on non-linguistic tests which make the same cognitive demands. It is performance on phonological tests which is systematically related to reading ability in the partial correlations analysis and the cross-lag analysis, and performance on these tests tends to load upon the same factors as are associated with reading ability in the factor analysis. As for the possibility that IQ is the primary factor which determines reading ability, Stanovich, Cunningham, and Freeman (1984b) have presented data which argue that phonological awareness is a more important factor in early reading ability than general IQ. Our own results are consistent with this fact and we note that the only component of the IQ test which successfully predicted reading ability was a vocabulary test which surely involved perception and retrieval of phonological structures.

In general, the results obtained with the non-linguistic controls indicate that the test battery can only predict reading ability to the extent that it places special demands on phonological skills. Some poor readers in the spring sample did encounter difficulty with the Corsi block test, but this result is not apparent in the fall sample, nor has it been obtained in at least one other study (Mann & Liberman, 1984). Also, Corsi-block performance has not been associated with reading ability in the second and fourth grades (Mann et al., in press). Perhaps a more interesting result concerns the test of environmental sound perception: On that test, some poor readers actually surpassed good readers. This result, together with the general pattern of our data, is consistent with the biological model of developmental dyslexia that was proposed by Geschwind and Galaburda (1985). According to their model, dyslexia arises out of an imbalance of hemispheric development which favors the right hemisphere to the detriment of the left hemisphere, an imbalance that gives rise to low scores on tests of left-hemisphere processing and superior scores on tests of right-hemisphere processing. Poor readers would be expected to excel on tests of right-hemisphere processing, and the perception of environmental sounds is well-known to recruit the right hemisphere (for references, see Brady et al., 1983), whereas they should fall behind good readers on language tests, as these recruit the left hemisphere.

The third and final point concerns the observation that the phonological skills of poor readers tend to resemble those of slightly younger children. This pattern of results supports a picture of "phonological delay" that is consonant with other suggestions in the literature on reading disability (for reviews, see Mann, 1986a; Mann & Liberman, 1984, Mann et al., in press). Rather than being deviant in their acquisition of phonological skills, poor readers in the early elementary grades may merely be maturing more slowly than other children of the same age. Our use of the word 'delayed' is not meant to imply that language skills of poor readers ultimately 'catch up' with those of their better reading peers. Indeed, current research is showing us that many of the special characteristics of disabled readers at this age appear to hold in adolescence (McKeever & van Deventer, 1975) and beyond (Jackson & McClelland, 1979; Read & Ruyter, 1985; Russell, 1982; Scarborough, 1984).

We therefore use the word 'delayed' to describe the salient characteristics of poor readers in the early elementary grades, and remain open to the possibility that a delay in early development can associate with inferior language skills at maturity, or even cause them.

For the beginning reader, language skills seem to be the most consistent area of delay, as non-linguistic abilities are considerably more variable and delays are less consistent. This possibility warrants careful consideration and further research, given the popular use of grade retention as a form of remediation. If confirmed, the notion of a language-specific delay implies that at least some poor readers could be capable of normal progress in math and geography as long as they are not hindered by their inferior reading skills. It also suggests that remediation in specific areas of school work will be more effective than repeating or delaying an entire grade.

In summary, this study of American children has illustrated that phonological deficiencies can be harbingers of reading failure. In the future it will be important to ask whether the relation between phonological skills and future reading ability generalizes to beginning readers in other countries which employ alphabetic orthographies. We will also enhance and extend this line of research by considering the determinants of reading ability in those countries where syllabaries and logographies are employed. Having begun to isolate the deficiencies which give rise to reading problems, we should contemplate the use of this valuable information, for it is not enough to determine which children are at "risk" for reading failure, we should also learn how to remediate their problems most effectively. Training procedures that can improve phonological awareness have been of interest in recent research and the results are quite promising (see Bradley and Bryant, 1985, for example). The time is ripe for considering whether training methods can improve working memory and lexical access as well.

ACKNOWLEDGMENT

The research described in this paper was supported by NICHD grant HD01994 to Haskins Laboratories, and by NICHD grant HD21182-01 to V. A. Mann at Bryn Mawr College. Preparation of this chapter was completed at the University of California, Irvine.

REFERENCES

Bradley, L., & Bryant, P. (1985). *Rhyme and reason in reading and spelling*. Ann Arbor: University of Michigan.

Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology, 35*, 345-367.

Blachman, B. A. (1983). Are we assessing the linguistic factors critical in early reading. *Annals of Dyslexia, 33*, 91-109.

Crowder, R. (1982). *The psychology of reading*. New York: Academic Press.

Denckla, M. B. (1973). Development of speed in repetitive and successive finger movements in normal children. *Developmental Medicine and Child Neurology, 15*, 635-645.

Geschwind, N., & Galaburda, A. M. (1987). *Cerebral lateralization*. Cambridge, MA: Bradford Books.

Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity*. New York: Harcourt, Brace & Jovanovich.

Jackson, M., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General, 108*, 151-181.

Katz, R. B. (1986). Phonological deficiencies in children with reading disability: Evidence from an object naming task. *Cognition, 22*, 225-257.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology, 32*, 474-484.

Liberman, I. Y. (1982). A language-oriented view of reading and its disabilities. In H. Mykelbust (Ed.), *Progress in learning disabilities* (Vol. 5). New York: Grune and Stratton.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. Kavanaugh & R. Venezsky (Eds.), *Orthography, reading and dyslexia* (pp. 137-154). London: University Park Press.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Mann, V. A. (1984). Longitudinal prediction and prevention of reading difficulty. *Annals of Dyslexia, 34*, 117-137.

Mann, V. A. (1986a). Why some children encounter reading problems: The contribution of difficulties with language processing and linguistic sophistication to early reading disability. In J. K. Torgesen & B. Y. Wong (Eds.), *Psychological and educational perspectives on learning disabilities* (pp. 133-159). New York: Academic Press.

Mann, V. A. (1986b). Phonological awareness: The role of reading experience. *Cognition, 21*, 65-92.

Mann, V. A., Tobin, P., & Wilson, R. (1987). Measuring phonological awareness through the invented spellings of kindergarten children. *Merrill Palmer Quarterly, 33*, 365 - 392.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities, 17*, 592-598.

Mann, V. A., Cowin, E., & Schoenheimer, J. (in press). Phonological processing, language comprehension and reading ability. Journal of Learning Disabilities.

Mann, V. A., Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language, 11*, 627-643.

McKeever, W. F., & van Deventer, A. D. (1975). Dyslexic adolescents: Evidence of impaired visual and auditory language processing associated with normal lateralization and visual responsivity. *Cortex, 11*, 361-378.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A., & McCutchen, D. (1982). Speech processes in reading. *Speech and Language: Advances in Basic Research and Practice, 7,* 237-269.

Read, C., & Ruyter, L. (1986). *Reading and Special Education, 6,* 8-17.

Russell, G. (1982). Impairment of phonetic reading in dyslexia and its persistence beyond childhood—research note. *Journal of Child Psychology and Child Psychiatry, 23,* 459-475 (b).

Scarborough, H. S. (1984). Continuity between childhood dyslexia and adult reading. *British Journal of Psychology,75,* 329-348.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorders: A modular approach. *Cognition, 21,* 139-168.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fisher, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Perception & Performance, 5,* 531-545.

Stanovich, K. E. (1982a). Individual differences in the cognitive processes of reading: I. Word decoding. *Journal of Learning Disabilities, 15,* 485-493.

Stanovich, K. E. (1982b). Individual differences in the cognitive processes of reading: II. Text-level processes. *Journal of Learning Disabilities,15,* 549-554.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38,* 175-190.

Stanovich, K. E., Cunningham, A. E., & Freeman, D. J. (1984). Intelligence, cognitive skills and early reading progress. *Reading Research Quarterly, 14,* 278-303.

Treiman, R., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In T. G. Waller & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 3). New York: Academic Press.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Review, 101,* 192-212.

Wolf, M., & Goodglass, H. (1986). Dyslexia, dysnomia and lexical retrieval: A longitudinal investigation. *Brain and Language, 28,* 159-168.

Woodcock, R. W. (1973). *Woodcock Reading Mastery Tests.* Circle Pines, MN: American Guidance Services.

FOOTNOTE

[†] Also, Department of Cognitive Sciences, University of California, Irvine.

Consideration of the Principles of the International Phonetic Alphabet*

Arthur S. Abramson†

As part of the preparation for the meeting on the revision of the International Phonetic Alphabet, I have been asked to write a short summary of my correspondence with a number of members of the Association on the principles on which the alphabet should be based. To this end, I have sought the views of the 34 respondents who gave first or second rank to this topic on the pre-registration form for the 1989 conference.

Unfortunately, by the time I was able to send out my letter with a few suggestive questions and the relevant passage from Ladefoged and Roach (1986: 25-26), only 13 scholars were able to send me their thoughts by the deadline or very shortly thereafter. They are: Michael Ashby, John Baldwin, J. C. Catford, Erik H. Erametsa, Eli Fischer-Jørgensen, Caroline G. Henton, Antti Iivonen, Klaus Kohler, Peter Ladefoged, Asher Laufer, André Martinet, Richard C. L. Matthews, and Francis Nolan. A few of them offered penetrating remarks on the topic as a whole; others commented briefly on most or all of the principles, and some chose one or two principles for discussion.

THE HISTORIC PRINCIPLES

For the convenience of the reader, the principles are reprinted here:

1. There should be a separate letter for each distinctive sound; that is, for each sound which—being used instead of another in the same language—can change the meaning of a word.
2. When any sound is found in several languages, the same sign should be used in all. This applies also to very similar shades of sound.
3. The alphabet should consist as much as possible of the ordinary letters of the Roman alphabet, as few new letters as possible being used.
4. In assigning values to the Roman letters, international usage should decide.
5. The new letters should be suggestive of the sounds they represent, by their resemblance to the old ones.
6. Diacritics should be avoided, being trying for the eyes and troublesome to read.

Haskins Laboratories SR-97/98

Status Report on Speech Research *1989*

SUMMARY OF THOUGHTS

Perhaps the six principles should be kept in print for their historical interest. A good way to do it might be to promulgate new principles, while preserving the old ones in a short statement of the history of the Association. In this connection, it should be borne in mind that an early goal was the fostering of good writing systems for the languages of the world. This may no longer be an important aim of the IPA.

Since Principle 1 implies the concept of the phoneme, we could consider expanding and clarifying it to make room for various theoretical approaches to the choice of phonologically relevant units. At the same time, perhaps it is not our business to go out of our way to accommodate various approaches to phonological theory or to take a particular stance; we are concerned, rather, with ways of representing graphically the human capacity for the production and perception of speech. In any event, to the extent that we *are* concerned with phonemic distinctiveness, do we want to obscure a basic relatedness—such as nasality or vowel length—between groups of phonemes by refusing to form symbols by combining a diacritic with several letters? (The latter point impinges upon Principle 6.) In addition, we surely need a large enough array of symbols to handle all the phonetic categories permitted by human physiological and psychoacoustic constraints, even if they have not all been found as yet in languages.

Principle 2 is going to have to be clarified. There is a sense in which the reader can accept it without quibbling. If certain sounds have the same structural status in a number of languages but differ qualitatively in a rather limited way, we may be content to assign the same symbols to them for texts in each of the languages; however, for comparative analysis and for language teaching, it may be useful to write them differently.

Principle 3 does not seem troublesome, although in practice some Greek letters, as well as a few others, are useful.

Opinions range widely on Principle 4, from rejection through skepticism to acceptance. We might ask whether there is really a reliable canon of international usage for Roman letters beyond correlation with rather gross phonetic categories. One interpretation offered is the notion that from nation to nation local orthographic biases might prevail. Another is that the phonological concept of markedness, together with universal occurrence, might guide us.

Principle 5 seems reasonable.

Principle 6, which seems to be in conflict with Principle 3, bothers many of us. While avoiding excessive diacritics, which are hard to print and read, we certainly want to have a number of them available. They can be very useful for symbolizing systemic relations and for the notation of fine detail in field work. Indeed, Principle 5 can sometimes best be carried out with the help of diacritics.

CONCLUSION

Given the important advances in our field since the days of the founders of the Association, the principles of the IPA need to be revised somewhat as part of our undertaking to improve the alphabet itself. The collection of responses underlying the present summary will, I think, furnish an excellent starting point for a fruitful discussion and good outcome in Kiel.

REFERENCE

Ladefoged, P., & Roach, P. (1986). Revising the International Phonetic Alphabet: A plan. *Journal of the International Phonetic Association, 16,* 22-29.

FOOTNOTE

[*] To appear in the *Journal of the International Phonetic Association, 18,* (1988).

[†] Also University of Connecticut, Storrs.

Language Specificity in Lexical Organization: Evidence from Deaf Signers' Lexical Organization of ASL and English*

Vicki L. Hanson[†] and Laurie B. Feldman

A sign decision task, in which deaf signers made a decision about the number of hands required to form a particular sign of American Sign Language (ASL), revealed significant facilitation by repetition among signs that share a base morpheme. A lexical decision task on English words revealed facilitation by repetition among words that share a base morpheme in English and ASL, but not among those that share a base morpheme only in ASL. This outcome occurred for both deaf and hearing subjects. Results are interpreted as evidence that the morphological principles of lexical organization observed in ASL do not extend to the organization of English for skilled deaf readers.

Repetition priming studies have demonstrated an appreciation of morphological structure by users of spoken languages. Specifically, studies have shown that when a target word follows the presentation of either an inflectionally or a derivationally related prime word, responses to the target are facilitated. For example, responses to the target word *sit* will be facilitated when preceded by the inflectionally related verb *sits* or by the derivationally related noun *sitter* (Fowler, Napps, & Feldman, 1985; Stanners, Neiser, Hernon, & Hall, 1979). Note that the triad *sit, sits*, and *sitter* share the base morpheme *sit*, but differ with respect to final morpheme (specifically Ø, -s, -er). Therefore, they represent a set of morphologically related words.

In contrast, no significant facilitation is obtained for targets following the presentation of orthographically related (but morphologically unrelated) items (Feldman, 1987; Feldman & Moskovljević, 1987; Murrell & Morton, 1974; Napps & Fowler, 1987). Thus, for example, there is no demonstrable facilitation for the target *sit* following the presentation of the word *site*. *Sit* and *site*, although formally (visually) similar, do not share a base morpheme. This outcome has been interpreted as evidence that formal similarity alone cannot account for the pattern of facilitation produced by the repetition priming paradigm.

Morphologically related words are obviously also related semantically and sometimes associatively. Facilitation due to semantic or associative priming has been experimentally demonstrated to persist only at short intervals between prime and target, specifically, lags of 1 or 2 items (Dannenbring & Briand, 1982; Henderson, Wallis, & Knight, 1984; Ratcliff, Hockley, & McKoon, 1985). In contrast, the studies reporting effects of morphological priming have found these effects to be more long lasting. In fact, these effects have been obtained at lags as long as 48 items (Fowler et al., 1985).

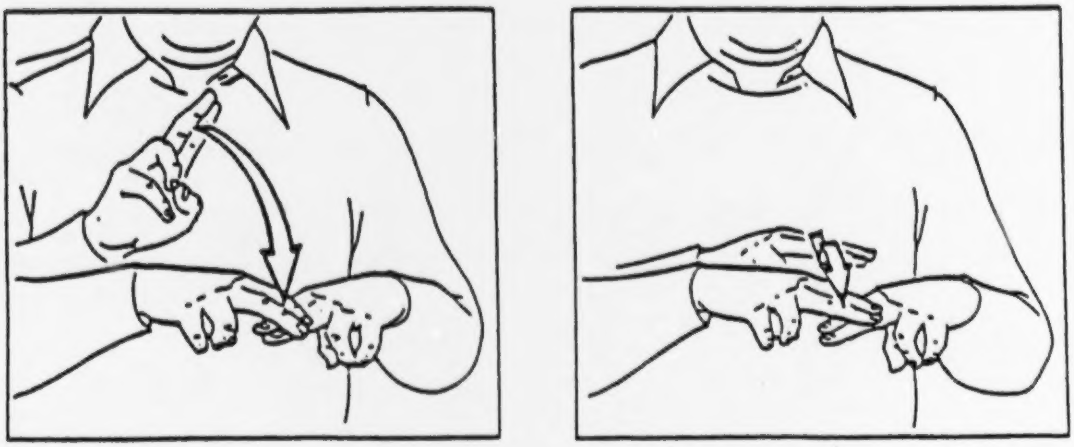
The finding that facilitation is produced under conditions of morphological priming has been interpreted as evidence that morphologically related words are stored close together in the internal lexicon. The particular form of this organization has been debated: it has variously been characterized as a base entry with associated tags for related words (Murrell & Morton, 1974; Stanners et al., 1979; Taft & Forster, 1975), as fully formed words that are linked to relatives (Lukatela, Gligorijević, Kostić, & Turvey, 1980) or as some combination of the two (Fowler et al., 1985). Despite discussion about how best to represent this lexical knowledge, the general claim that facilitation in the repetition priming task reflects a morphological principle of organization within the lexicon is widely accepted.

In the present study, we examine lexical organization by deaf bilinguals of American Sign Language (ASL) and English. We are interested in whether the lexical organizations of the two languages are the same, or whether there is language specificity, such that signs of ASL are organized in accord with relationships inherent in ASL morphology, while words of English are organized in accord with relationships inherent in English morphology.

ASL is the common form of communication used by members of deaf communities across the United States and parts of Canada. Deaf children may acquire ASL as a first language either through contact with the deaf community or from deaf parents. It is a primary visual-gestural system, autonomous from other languages. It is not based on, nor derived from, any form of English. Contrast, for example, lexical structuring for SIT in English and ASL. In English, base lexical items are composed of *sequentially* structured phonemic segments (e.g., /s/ /I/ /t/). In ASL, however, base lexical items are composed of the *co-occurring* formational parameters of handshape, movement, and place of articulation (Stokoe, Casterline, & Croneberg, 1965). For example, in ASL the sign for SIT is produced, as shown in Figure 1, with H-handshapes, the place of articulation in neutral space in front of the signer, and the movement of the dominant hand being a single sweeping motion contacting the non-dominant hand.

One of the intriguing aspects of ASL structuring is that, in contrast to English in which morphological processes generally operate by the affixing of a sequence of morphemes (composed of phonemic segments), for example, -ing, -ed, -er, to base morphemes, morphological processes in ASL may involve embedding the base morpheme within a specific pattern of movement, such that the base morpheme of the sign and the inflectional or derivational morphemes of the sign co-occur. As an example, one of the many morphological relationships marked in ASL is a Noun-Verb distinction. According to the linguistic analysis of derivationally related Noun-Verb pairs such as CHAIR-SIT (Supalla & Newport, 1978), the sign for CHAIR and the sign for SIT share an underlying representation (base morpheme) in which information is stored about such features as handshape, place of articulation, movement and direction of movement. The surface forms of these signs are presumed to be derived by modulations applied to this common base. As shown in Figure 1, the verb SIT has a single, relatively free movement, while the noun CHAIR has the same movement in a duplicated and restrained (or smaller) form.

While morphological processes of ASL have been linguistically described, it is unclear under what conditions, if any, these relationships influence signers' lexical organization. Earlier work has shown that deaf signers decompose morphologically complex signs into their base and morphological marker when remembering single signs and signed sentences (Hanson & Bellugi, 1982; Poizner, Newkirk, Bellugi, & Klima, 1981). For example, in a short-term memory task, Poizner et al. (1981) presented short lists of inflected signs to deaf subjects. In recalling these lists of signs, the subjects would sometimes recall a base sign with an inflection from a different item in the test. For example, when presented with lists of signs containing the morphologically complex signs corresponding to the English phrases TAKE-ADVANTAGE-OF-THEM and PAY-ME, a typical error was recalling the signs TAKE-ADVANTAGE-OF-ME and PAY-THEM. This outcome was taken as evidence that the signers were remembering the base entry and the inflectional tags separately. This result, however, does not address the issue of how these base signs (or composites of signs) are organized in the signer's lexicon.



SIT

CHAIR

Figure 1. An example of a derivationally related Noun-Verb pair in ASL: the signs SIT and CHAIR. (From Baker, C., Cokely, D. (1980). *American Sign Language: A teacher's resource text on grammar and culture.* Silver Spring, MD: T. J. Publishers, p. 105.)

If, as occurs in experimental tasks conducted with English materials, the underlying morphological structure is abstracted from signs, then we would expect the lexical organization of ASL to reflect morphological relationships. Experiment 1 examines this question. As the test case, we have chosen to examine the case of morphologically related noun-verb pairs because this morphological relationship has been well studied (Supalla & Newport, 1978). Because the morphological structuring in ASL, as in the case of the noun-verb pairs used in Experiments 1 and 2, represents a departure from how morphological relationships are marked in English, we are interested here in whether a morphological distinction that is marked by a global feature (in this case, movement) can be abstracted in a manner functionally equivalent to the way a user of English abstracts the underlying morphological structure of that language. Given that ASL has been shown in the linguistic (see Padden, 1987, for a discussion) and psycholinguistic (see, for example, Poizner, Bellugi, & Klima, 1987) research to date to operate according to the same principles as other natural languages, we expect that evidence of lexical organization based on ASL morphology will be obtained. If so, then we can proceed to the present question of primary interest, which is whether deaf readers' lexical organization of English reflects relationships inherent to ASL.

In order to examine how morphological relatives are stored in the ASL lexicon of deaf signers, a task requiring subjects to respond in a "sign mode" was necessary. For this purpose we developed a *sign decision task*. In this task, as in most repetition priming studies, subjects were presented with printed target and prime items. Rather than deciding if these items were words in English, however, in the sign decision task they were required to indicate the number of hands that they use to produce the sign for each word. Therefore, as each English word was presented, they had to think of the sign for the word, and then make a decision *based on the physical form of that sign*. Note that it would not have been possible to perform this task based simply on the English forms of the words. Thus, the sign decision task is not an English reading task; it is a task that uses the subjects' knowledge of ASL. Nevertheless, as in a lexical decision task, items were presented individually and the subjects were asked to respond as quickly and as accurately as possible and reaction time was measured.

The analysis of subjects' accuracy in the sign decision task is not as straightforward as the analysis of a lexical decision task. There are, for example, dialectal differences that cause signers from different parts of the country to produce different signs for a given English word. For this reason, following the sign decision task subjects were given an unspeeded paper and pencil version of the task in which they were again asked to indicate, this time in writing, the number of hands that they used to produce the sign for each word listed. The correctness of each subject's answers in the timed sign decision task was then determined relative to each subject's own stated sign productions.

Assuming that sign priming can be demonstrated for deaf subjects, we can then ask whether deaf bilinguals also organize their English lexicon according to morphological relatedness among signs. In Experiment 2, we investigate lexical organization of English by deaf bilinguals. Experiments 1 and 2 each used a set of words that are morphologically related only in ASL and a set that are morphologically related in both ASL and English. The same set of ASL Only words were used in the two experiments. The words in the ASL and English sets in the two experiments were not identical, although there was substantial overlap in the two sets.

The ASL Only stimuli were derivationally related Noun-Verb pairs that shared a base morpheme and were morphologically distinguished by a change in movement superimposed throughout the production of the sign. Examples are EAT/FOOD, SAIL/BOAT, DRIVE/CAR, LOCK/KEY, and SHOOT/GUN. Note that these items are also associatively related.

The English + ASL set of stimuli were words that are morphologically related in *both* English and ASL. Examples of pairs from this condition are JUMP/JUMPING, TEACHER/TEACH, and CREATING/CREATE. For these pairs, the morphological relationship in ASL was sometimes structurally different than that of the ASL Only set. For example, in pairs such as TEACHER/TEACH, the morphological marker used sequential rather than simultaneous structuring. Note that although these pairs are semantically related, they are not associatively related in that when given the word *teacher* subjects are much more likely to free associate *student* than they are to free associate *teach*.

To guard against influences of associative and semantic priming, the present study used an average lag of 10 items, a lag that is too great to produce such priming

(Dannenbring & Briand, 1982; Henderson et al., 1984; Ratcliff et al., 1985). Consequently, any facilitation obtained in the present priming study can be convincingly attributed to morphological, not associative or semantic, priming.

To summarize, in Experiment 1 we examined whether signers' lexical organization of ASL is based on relationships inherent to ASL morphology. For this purpose, the sign decision task was developed. The experimental materials consisted of target words and their morphologically related primes. The analyses focused on decision latencies and errors to targets facilitated by the prior presentation of a related prime. The purpose was to investigate morphological priming under conditions in which subjects are responding based on their knowledge of ASL. In Experiment 2, the same materials were presented to deaf and hearing subjects in a lexical decision task of English. Analogous to Experiment 1, the analyses focused on decision latencies and errors to targets preceded by morphologically related primes. In Experiment 2, subjects responded based on their knowledge of English. The purpose of Experiment 2 was, therefore, to examine whether deaf signers' organization of English words is based on relationships inherent to English or to ASL morphology. Stated differently, in Experiment 2 the role of the morphological structure of ASL in the lexical organization of English was investigated.

EXPERIMENT 1

Method

Subjects

All were prelingually and profoundly deaf, with hearing losses of 85 dB or greater in the better ear. Twelve of these subjects were native signers of ASL and another two came from deaf families (i.e., older deaf siblings and deaf grandparents). The remaining subjects had acquired ASL before they entered school and reported ASL to be their preferred means of communication with deaf and hearing friends. All were paid volunteers for this 15 min experiment. As measured by comprehension subtest of the *Gates-MacGinitie Reading Tests* (1978, Level F, Form 2), their median reading grade level of English was 7.4 (N16, Range 12.9 to 3.3).

Stimuli

The experimental materials consisted of 28 pairs composed of target word and a morphologically related prime. Two sets of materials, distinguished by the language in which they were morphologically related, were created: ASL Only, and English + ASL. Each set included 14 pairs.

In the ASL Only set, all paired items were derivationally related noun-verb pairs. For example, the target word CHAIR was paired with the derivationally related prime word SIT. As described earlier and shown in Figure 1, CHAIR and SIT share the same handshape and place of articulation, but differ in their movement. SIT has the single motion of a verb, while CHAIR has the more constrained double movement of a noun. These word pairs were taken from the list of Noun-Verb pairs in Supalla and Newport (1978). The experimental set represents all the words that fit within the following experimental constraints: 1) the items in the noun-verb pairs correspond to different English words (thus, excluding pairs such as SKI/SKI), 2) there was on unambiguous English interpretation for each item in a noun-verb pair (thus, excluding pairs such as CIGARETTE/SMOKE in which SMOKE could be correctly translated as the verb, or, for this study, incorrectly translated as the noun), and 3) each item does not require multiple English words to translate the ASL sign (thus excluding pairs such as BABY/ROCK-BABY-IN-ARMS).[1]

In the English + ASL set, base words served as targets, while words inflectionally and derivationally related to these base words served as primes. The pairs were subject to the following constraints: in both target and prime, the stem undergoes 1) no change in spelling, and 2) no change in pronunciation in English. For example, in this set the target word "wish" was paired with the morphologically related prime word "wishing." A complete list of the stimuli is given in the Appendix.

Two test orders were created. Each included equal numbers of primed and unprimed trials for both ASL and English + ASL items. The lists were constructed so that targets preceded by morphological primes in one list were preceded by unrelated words in the second list and vice versa. Each subject saw only one test order, thus responding to each target only once, in either a primed or an unprimed context.

Each list consisted of one block of 118 trials, presented in a fixed order. Prime and target items were separated by an average of 10 intervening items. (The range was 8 to 12). Filler items were included to preserve appropriate lags. These filler items were words, chosen to maintain approximately equal numbers of one- and two-handed sign responses.

Procedure

The start of each trial was signaled by a 250 ms fixation point (+) presented in the center of a CRT display, followed by a 250-ms blank interval. The letter string was then presented in the center of the CRT display. The stimulus remained in view either until the subject pressed a response key or until 5 s had elapsed. Feedback was given on each trial. The feedback was the subject's RT (in ms) for that trial. If the subject had failed to respond within the 5 s time limit, the words TOO SLOW appeared as feedback. The feedback, displayed for 250 ms, was centered six lines below the fixation point. Following the message, there was a one second interval before the start of the next trial. Prior to testing with the experimental stimuli, subjects received one practice block of 15 trials.

The subjects were instructed that on each trial they would see a letter string and that their task was to decide, as quickly and as accurately as possible, whether they made the sign for that word with one or two hands. There were two response keys; one was labeled '1' and the other was labeled '2'. They were instructed to sit with their index fingers resting one on each key and to respond as quickly as possible. The subjects were also informed about the nature of the feedback. The experimenter was a deaf native signer, a former Gallaudet student, who communicated with the subjects by signing.

Following completion of the sign decision task, subjects were given a paper on which 49 words were typed from the test order that they had been presented. These 49 words consisted of the 28 target items (14 from each set), the 14 morphologically related primes that had occurred in the test order for that subject, and 7 of the 14 unrelated "primes" from that list order. Each word was typed, in uppercase letters, with a line preceding it. The following instructions were typed at the top of the paper:

> Below are listed the words you just saw in the sign decision task on the computer. Different people sometimes sign a few of these words in different ways. For each word, please indicate if YOU sign the word with one hand or two hands. For example:
>
> 2 BABY
>
> 1 EYE

Results

Subjects' own judgments in the paper and pencil task were used to determine the correct answer for each item in the sign decision task. Thus, the correct response in the sign decision task was conditionalized on each subject's answers in the paper and pencil task. RTs were stabilized by eliminating responses whose RTs differed by more than two standard deviations from a subject's mean. The resulting mean RTs and mean percentage errors are shown in Table 1.

TABLE 1: The mean RT (in ms) for the deaf subjects in the sign decision task of Experiment 1. The mean percentage errors are given in parentheses.

	English + ASL	ASL Only
Unprimed	1085 (13.5)	1163 (12.8)
Primed	1013 (12.8)	1038 (9.0)

An analysis of variance was performed on the latencies for the within-subject variables of prime condition (primed vs. unprimed) and language of morphological relation (English + ASL vs. ASL Only). The analysis for RTs revealed a significant main effect of prime condition, $F(1,18) = 5.76$, $MSe = 31978.23$, $p < .03$. On an items analysis, with language of morphological relation being treated as a between-items factor, the main effect of prime condition was also significant, $F(1,26) = 14.08$, $MSe = 10835.81$, $p < .02$. There was no effect of language of morphological relation, $F(1,18) = 2.90$, $MSe = 17249.42$, $p > .10$ for subjects; $F < 1$ for items. The interaction of the two variables was not significant, $Fs < 1$ for both subjects and items. No significant effects were revealed in the error analyses, all $Fs < 1$.

Finally a t-test revealed that the 125 ms facilitation with the ASL Only set was significant, $t(18) = 2.15$, $p < .05$ for subjects; $t(13) = 3.02$, $p < .01$ for items. This outcome is critically important in the interpretation of both this experiment and Experiment 2 and will be treated in the Discussion of Experiment 2.

Discussion

In Experiment 1, significant facilitation for morphologically related items was obtained with materials of both the English + ASL set and the ASL Only set. Interpretation of the facilitation for English + ASL is ambiguous as to whether lexical organization is based on relationships derived from English or from ASL. However, the fact that comparable facilitation was obtained for the ASL Only set indicates that lexical organization can be based on principles of ASL. The obtained facilitation for the ASL Only stimuli cannot be explained as being due to associative priming, as the lags introduced in the present study were too great to permit such priming (Ratcliff et al., 1985; see also Dannenbring & Briand, 1982). This pattern of results suggests that users of ASL tend to organize their lexicons according to morphological relationships.

It is worthy of note that the results were obtained using prime-target pairs in which the morphological relationship was indicated by a process quite different from the morphological processes that generally operate in English. In Experiment 1, the noun-verb distinction of the pairs was indicated by a change in the movement throughout the sign production.

Arguably, the effect here could be visual, based on the visual similarity of prime and target signs for the pairs in the ASL Only set. This argument is difficult to support, however. This argument for facilitation based on the visual similarity of signs would be analogous to an argument based on the visual (orthographic) similarity of printed words or the phonetic similarity of spoken words. Evidence from different paradigms (e.g., lexical decision, recognition) has established, however, that formal similarity, in the absence of a shared base morpheme, does not facilitate responding for any but the briefest intervals (Feldman & Moskovljević, 1987; Hanson & Wilkenfeld, 1985; Kirsner, Dunn, & Standen, 1987; Murrell & Morton, 1974; Napps & Fowler, 1987). As noted earlier, the linguistic and psychological evidence give no reason to suggest that ASL signs would be processed differently (e.g., Padden, 1987; Poizner et al., 1987). For example, in early research on ASL there was much discussion of the role that iconic aspects of signs would play in signers' processing of signs. This turned out not to be relevant. Signers, in short-term memory situations, have been found to process signs as linguistic events, not as iconic "pictures" (Poizner, Bellugi, & Tweney, 1981) or as unanalyzed "wholes" (Hanson & Bellugi, 1982; Poizner et al., 1981). There is little reason, therefore, to expect that it is the visual/motoric similarity between the signs for words in the ASL Only set that is responsible for producing the response facilitation observed in Experiment 1.

Given that the ASL lexicon of signers suggests a principle of organization based on the morphology of ASL, how is their lexicon of English organized? In Experiment 2 we investigate whether the English lexicon is influenced by the morphological relationships of ASL.

From the bilingual literature, there is reason to expect that ASL organization would not influence the organization of a signer's English lexicon. For example, in studies that have looked at whether a word in one language activates its translation in another language, influences of one language on the processing of the other have not been obtained, provided that the experiment has not been set up to encourage such translation (Kirsner, Brown, Abrol, Chadha, & Sharma, 1980; Kirsner, Smith, Lockhart, King, & Jain, 1984; Scarborough, Gerard, & Cortese, 1984; Watkins & Peynircioğlu, 1983; see also Beauvillain & Grainger, 1987). The lexical decision task of Experiment 2 is not one that requires translation from one language into another. That is, it does not test whether the printed words SIT and CHAIR are translated into their corresponding ASL signs.

Another finding from the bilingual literature also would lead us to expect that ASL principles would not guide the organization of the English lexicon for the subjects used in the present study. Research with bilinguals has found evidence for translation into the first language only for readers who have poor mastery of the second language (Kroll & Borning, 1987). The deaf subjects of the present study have a fairly good mastery of their second language, English. As we will see, the deaf subjects in Experiment 2 were reading, on average, at the tenth grade level.

The present study differs from these previous bilingual studies, however, in terms of the radically different structuring of the two languages studied. Consider the difference between the case of a bilingual of two spoken languages and of the ASL/English bilingual. For bilinguals of two spoken languages, the building blocks of the two languages will be similar. That is, the phonological and morphological segments in both languages are articulated by gestures of the vocal tract. This is true whether the two languages are similar, such as English and Spanish, or relatively different, such as English and Chinese. In contrast, the building blocks of ASL and English are quite different. Whereas in ASL the phonological[2] and morphological elements are *visible* gestures articulated by the hands, face, and body of the signer, in English these elements are *audible* gestures articulated by the vocal tract of the speaker. Moreover, not only do the building blocks differ for ASL and English, but in learning English, its building blocks are not readily accessible to the deaf individual. If tests of the speaking and lipreading skills of deaf individuals are used as an indicator, it is apparent that the underlying phonemic representation of English words is not always well-mastered by deaf individuals (e.g., Smith, 1975).

Accordingly, the differences in processing in sign and speech may be so great, that for the deaf individual who has ASL as a first language and has established mental processing based on sign language parameters, language operations will continue to be carried out in ASL mode—regardless of the language of input. This is the primary language notion of Shand (Shand, 1982; Shand & Klima, 1981). This notion would predict that English words would be translated into their corresponding ASL signs, regardless of task requirements. As a consequence, the lexical organization revealed in an English task would be expected to reflect ASL organizational principles.

In summary, an argument can be made from the existing literature to support either the presence or the absence of an influence of ASL lexical organization when reading English. In Experiment 2, we sought to resolve this issue. In this endeavor, we employed a standard lexical decision (word/nonword) task in order to ask whether deaf signers exploit morphological relationships among ASL words while reading English materials. Two groups of subjects were tested. There was a group of deaf college students who reported ASL to be their first language. In addition, there was a control group of hearing college students. Based on the English repetition priming literature, we would expect hearing subjects to show facilitation to target words primed by morphological relatives in the English + ASL set, since both inflectional and derivational relationships have been shown to facilitate responding (Fowler et al., 1985; Stanners et al., 1979). Because only an English relationship can benefit the hearing subjects, morphological priming should be absent for items that were morphologically related only in ASL (*viz.*, the ASL Only set). Due to the rapid decay of associative priming, any associative relatedness between pairs in the ASL Only set would not facilitate responding under the lag conditions of Experiment 2. Of particular interest is whether the deaf subjects will show significant priming in both the ASL Only and the English + ASL sets, or only in the latter set. Experiment 1 demonstrated significant priming in both sets. If, due to experiences in sign, deaf signers organize their English lexicons along the same morphological principles that apply to ASL, then we would expect to see facilitation in English reflecting morphological relationships of ASL. Specifically, both stimulus sets should show facilitation. By contrast, if deaf readers maintain distinct morphological principles in their English and ASL lexicons then words that are related in their ASL lexicon need not be linked in their English lexicon. In this case, we would expect to see facilitation restricted to the English + ASL set.

EXPERIMENT 2

Method

Subjects

The deaf subjects for this study were 19 paid volunteers from Gallaudet University. All were prelingually deaf, with onset of deafness occurring prior to one year of age. Seventeen of these subjects had deaf parents; the other two subjects had learned to sign before entering school. Further, all sustained a profound hearing loss of 85 dB or greater, pure tone average. Available reading achievement test scores, measured on the comprehension subtest of the *Gates-MacGinitie Reading Tests* (1978, Level F, Form 2), indicated a median reading level of grade 10.2 (Range 12.9 to 6.0, N 14). This sample group of subjects therefore represented a relatively skilled subset of the population of deaf readers.

The hearing subjects for this study were 20 undergraduates at the University of Delaware who participated in partial fulfillment of the research requirements in an introductory Psychology course. All were native speakers of English with normal or corrected-to-normal vision.

Stimuli

The experimental materials overlapped those of Experiment 1. The ASL Only set contained the same 14 prime-target pairs used in Experiment 1. The 14 prime-target pairs in the English + ASL set differed from those of Experiment 1, but were selected according to the same criteria.[3] A complete list of the stimuli is given in the Appendix.

The experimental materials consisted of two lists of 118 stimuli. Both lists were composed of equal numbers of words and pseudowords. Each target (word or pseudoword) was separated from its prime by an average of 10 intervening items (range 7 to 13). Half of the items in each list were preceded by morphologically related primes and half were not. Across lists, each target was preceded by its related prime.[4] Each subject saw only one of the experimental lists, which assured that the target word was viewed only once by a subject, in either a primed or an unprimed context and that all subjects viewed both primed and unprimed targets.

Pseudowords were created by changing one or two letters in the words. Consequently, for English + ASL pseudowords, prime and target were physically similar (e.g., LELPER/LELP for HELPER/HELP), whereas for ASL Only pseudowords, there was no similarity of prime and target (e.g., GOOM/OIT for FOOD/EAT.)

Procedure

The procedure was identical to that of Experiment 1, except for instructions as to task and some change in feedback. The feedback was the subject's RT (in ms) for that trial which, if the subject had made an error on that trial, was preceded by a minus sign.

The subjects were instructed that on each trial they would see a letter string and that their task was to decide, as quickly and as accurately as possible, whether or not the string was a real English word. The two responses were associated with different keys; the right-hand one was labeled YES and the left-hand one was labeled NO. The subjects were instructed to sit with their index fingers resting one on each key and to respond as quickly as possible. They were also informed about the nature of the feedback. For the deaf subjects, the experimenter was a deaf native signer, a former Gallaudet student, who communicated with the subjects by signing. For the hearing subjects, the experimenter was an undergraduate at the University of Delaware. Prior to testing with the experimental stimuli, subjects received one practice block of 15 trials. The total duration of the experiment was about 15 min.

Results

A preliminary items analysis indicated that the target word STRUM was missed by the hearing subjects on 80% of its presentations and by deaf subjects on 100% of its presentations. For that reason, this item was eliminated from all analyses for both the deaf and hearing subjects. As in Experiment 1, RTs that deviated more than two standard deviations from the cell means were eliminated from analysis. The resulting mean correct RTs and the mean percentage errors are shown in Table 2. There were no significant interactions involving subject group in any of the analyses reported below, all $ps > .05$. Thus, the same pattern of results was obtained for both the the hearing and the deaf subjects.

TABLE 2: The mean RT (in ms) to words and nonwords for the hearing and deaf subjects in the lexical decision task of Experiment 2. The mean percentage errors are given in parentheses.

	English + ASL		ASL Only	
	Hearing	Deaf	Hearing	Deaf
WORDS				
Unprimed	525 (.7)	519 (2.2)	546 (2.4)	558 (2.2)
Primed	484 (.7)	479 (.0)	552 (3.4)	550 (.7)
NONWORDS				
Unprimed	635 (15.3)	667 (4.1)	608 (2.5)	621 (2.7)
Primed	627 (10.0)	627 (10.7)	632 (9.9)	633 (11.7)

For word latencies, an analysis of variance on the within-subjects factors of prime condition (unprimed vs. primed) X language of morphological relation (English + ASL vs. ASL Only) and the between-subjects factor of group (deaf vs. hearing) indicated a significant interaction of condition by language of morphological relation, $F(1,37) = 6.91$, $MSe = 2143.20$, $p < .02$. This significant interaction also generalized across stimuli, as indicated by the items analysis with language as a between-items factor, $F(1,48) = 5.61$, $MSe = 1366.53$, $p < .03$.

Separate post hoc analyses on condition by group for each language revealed significant facilitation for English + ASL materials, $F(1,37) = 14.87$, $MSe = 2152.03$, $p < .001$, for subjects; $F(1,26) = 13.23$, $MSe = 1629.68$, $p < .002$ for items, but not for ASL Only materials, $Fs < 1$ for both subjects and items analyses. There was no interaction with subject group in either of these post hoc tests (both $Fs < 1$). Most critically, while the deaf subjects in Experiment 1 had shown a significant 125 ms facilitation with the ASL Only set, in Experiment 2, the deaf subjects showed only an 9 ms nonsignificant decrease in RTs, $ts < 1$ for both subjects and items. Thus, for both subject groups, significant facilitation was obtained following English + ASL morphological primes, but not following ASL morphological primes.

Finally, the overall ANOVA on RTs also revealed significant main effects of prime condition, $F(1,37) = 10.54$, $MSe = 1636.48$, $p < .005$ for subjects; $F(1,48) = 9.17$, $MSe = 1366.53$, $p < .005$ for items, and language of morphological relation, $F(1,37) = 46.24$, $MSe = 2071.31$, $p < .001$ for subjects; $F(1,48) = 15.96$, $MSe = 3831.47$, $p < .001$ for items. The effect of condition reflected faster RTs for primed than unprimed trials, and the effect of language reflected faster RTs for the trials in the English + ASL stimulus set.

An analysis of the errors on words indicated only a significant effect of language in the subjects analysis, $F(1,37) = 4.34$, $MSe = 14.81$, $p < .05$, reflecting somewhat fewer errors on the words used in the English + ASL set than the ASL Only set. This effect was only marginally significant in the items analysis, $F(1,48) = 3.64$, $MSe = 19.79$, $p < .07$.

In studies of repetition priming, the data from the nonword trials can be used as an index of whether the obtained facilitation is due to an experiment-specific (episodic) effect or to a lexical effect (Feustel, Shiffrin, & Salasoo, 1983; Fowler et al., 1985). Nonword facilitation is generally not considered to be lexical in origin. If significant facilitation is obtained in the nonword data, then episodic, rather than lexical, influences might also contribute to the outcome observed with words. The mean correct RTs and mean percentage errors for the nonwords are shown in Table 2. In the analysis of nonword RTs, the only significant finding was an interaction of prime condition by language of morphological relation in the subjects analysis, $F(1,37) = 5.59$, $MSe = 3049.80$, $p < .05$. This effect failed, however, to reach significance in the items analysis, $F(1,52) = .16$, $MSe = 1912.77$, indicating that the effect in the subjects analysis was due to a couple of problematic stimuli. In the analysis of the nonword error data, there was a significant effect of prime condition, $F(1,37) = 5.73$, $MSe = 135.21$, $p < .05$, and a significant interaction of prime condition by language of morphological relation, $F(1,37) = 4.25$, $MSe = 132.05$, $p < .05$, neither of which reached statistical significance in the items analysis, $Fs < 1$ for both. The results of the nonword data, therefore, give little reason to infer that episodic influences contribute to the pattern of results obtained for the word data.

Discussion

The same pattern of results was observed for the hearing subjects as for the deaf subjects in the English lexical decision task. That is, there was an interaction of prime condition by language of morphological relation such that significant facilitation was obtained only for prime-target pairs that were related in English. The significant facilitation with the English + ASL materials replicates earlier work by Hanson and Wilkenfeld (1985), namely, significant facilitation for targets preceded by primes that were morphologically related in both languages. The finding of interest in the present experiment is the absence of facilitation on trials in which ASL provided the only relationship of prime to target. This pattern of results suggests that the lexical organization of English materials by deaf readers is based on morphological relationships that are evident in English, not on those present in ASL. Thus, the lexical organization of these signers tends to be language specific, in that processing of ASL involves a structuring based on principles of ASL morphology, and processing of English involves a structuring based on English morphology.

The subjects' better performance on the English + ASL set than on the ASL Only set, as indicated by faster RTs and fewer errors, cannot be interpreted as reflecting an inherent difference in cognitive processing for English and ASL. Rather, this effect simply reflects subjects' differential familiarity with words in the two sets. During stimulus creation, these two sets were not equated for word frequency or familiarity.

GENERAL DISCUSSION

Important to our interpretation of language specificity in lexical organization are the different patterns of facilitation of the ASL Only morphologically related prime-target pairs in Experiments 1 and 2. In a task that required decisions on words of English (*viz.*, the lexical decision task of Experiment 2), the deaf subjects were not facilitated in their responding to pairs that were morphologically related only in ASL. In a task that required decisions on signs of ASL, however (*viz.*, the sign decision task of Experiment 1), the deaf subjects' responses were facilitated when prime-target pairs were related in ASL. To be able to conclude that the obtained facilitation reflects the ASL relatedness of the noun-verb pairs, it is necessary to rule out possible explanations of this effect that attribute it to the associative relatedness of the signs in the ASL Only set or, alternatively, due to the visual/motoric similarity of those signs.

As the pairs in the ASL Only set were associatively related, it bears considering whether the facilitation in Experiment 1 could be attributed to associative relatedness. To begin, we can use the pattern of facilitation across experiments to dismiss an account based on associative priming. In the ASL Only set, response facilitation occurred in the sign decision task of Experiment 1, in which responses were based on signs of ASL, but not in the lexical decision task of Experiment 2, in which responses were based on words of English. Had associative factors been responsible for the outcomes in Experiment 1, they should have been evident in Experiment 2 as well. It could be argued that such reasoning across experiments does not apply in the present case due to the fact that different tasks were used in the two experiments. Although there is evidence that some priming tasks produce associative priming while others do not (Henik, Friedrich, & Kellogg, 1983; Smith, Theodor, & Franklin, 1983), priming effects have been obtained only over lags of 1 or 2 items (Henik et al., 1983; Meyer, Schvaneveldt, & Ruddy, 1975; Ratcliff et al., 1985; Rugg, 1987; Smith et al., 1983; see also Dannenbring & Briand, 1982) even with tasks favorable to associative priming. The lag of 10 items in Experiment 1 is clearly beyond that at which associative priming can be expected to persist.

As discussed above, an account of Experiment 1 based on formal priming (i.e., physical similarity) is also inadequate. In consideration of the difficulty in demonstrating facilitation based on the formal similarity of prime and target in several lexical decision and recognition tasks at comparable (or longer) lags (Feldman & Moskovljević, 1987 Hanson & Wilkenfeld, 1985; Murrell & Morton, 1974) or even at shorter lags (Napps & Fowler, 1987), we conclude that formal similarity cannot govern the outcome in the sign decision task of Experiment 1.

Thus, the facilitation on the prime-target pairs that were morphologically related only in ASL in Experiment 1 can be attributed to lexical organization based on ASL. The absence of facilitation in the English lexical decision task for deaf subjects on the prime-target pairs that were morphologically related only in ASL suggests that the English lexicons of these signers is not organized according to relationships inherent to ASL. How is their English lexicon organized? We know from previous research that the English lexicons of deaf skilled readers are not organized visually/orthographically (Hanson & Wilkenfeld, 1985). Because the evidence in the present study indicates that it is not organized along principles of ASL morphology, the obtained facilitation in the present experiment and the earlier work on repetition priming with deaf readers (Hanson & Wilkenfeld, 1985) suggests that for deaf skilled readers the English lexicon is organized along a principle of English morphology.

The present evidence that principles of ASL morphology influence how deaf skilled readers process ASL signs but not how they process individual words of English is inconsistent with the primary language notion suggested by Shand (Shand, 1982; Shand & Klima, 1981). Although the exact meaning of "primary language" as used by Shand has always been vague, the basic assumption is that the language processing of deaf signers will be carried out in sign, regardless of whether the language input is sign or print. This notion clearly predicts, therefore, that deaf signers will show a tendency to process printed English words in terms of their sign equivalents.

In understanding the results of the present study, it is necessary to recall that the subjects of the study were quite skilled deaf readers. As noted earlier, they may not be representative of all deaf signers in that the average of their reading scores was exceptionally high. This leaves open the possibility that less skilled deaf readers might perform differently than the subjects who participated in the present study. It is still possible, therefore, that the primary language notion may better describe the English processing of less skilled deaf readers.

In conclusion, the outcome of Experiment 2 revealed no evidence that the lexical organization of English is influenced by characteristics of ASL. It should be recalled, however, that in Experiment 1 evidence *was* obtained for lexical organization based on ASL morphology in a task that required a decision based on signs of ASL. This pattern of results suggests that deaf skilled readers use morphological principles of ASL in their lexical organization of ASL, but that their lexical organization of English does not reflect morphological characteristics of ASL. In sum, lexical organization is language specific.

ACKNOWLEDGMENT

This research was supported by Grant NS-18010 from the National Institute of Neurological and Communicative Disorders and Stroke and by Grant HD-01994 from the National Institute of Child Health and Human Development. In supporting this work, we are grateful to Drs. Michael Karchmer and Patrick Cox at Gallaudet University. We wish to thank Carol Padden and Carol Fowler for their very constructive comments on earlier drafts of this manuscript, and to thank the reviewers for many helpful suggestions. We also wish to thank Deborah Kuglitsch and Terry Ullery for helping to test the subjects. This research was presented at the meeting of the Psychonomic Society, Seattle, Washington, November, 1987.

APPENDIX

Experiment 1

English + ASL		ASL Only	
Target	Morphological Prime	Target	Morphological Prime
WISH	WISHING	EAT	FOOD
CRASH	CRASHED	SAIL	BOAT
WALK	WALKED	SHOOT	GUN
CREATE	CREATING	TELEPHONE	CALL
HELP	HELPER	BROOM	SWEEP
COMPLAIN	COMPLAINING	WAITRESS	SERVE
DEMAND	DEMANDED	DRIVE	CAR
RELAX	RELAXED	CHAIR	SIT
JUMP	JUMPING	LOCK	KEY
KILL	KILLER	SCISSORS	CUT
TEACH	TEACHER	ICE CREAM	LICK
BORROW	BORROWED	SHOVEL	DIG
RIDE	RIDER	LAWNMOWER	MOW
LOVE	LOVER	STRUM	GUITAR

Experiment 2

English + ASL		ASL Only	
Target	Morphological Prime	Target	Morphological Prime
HELP	HELPER	EAT	FOOD
TURN	TURNING	SAIL	BOAT
CRASH	CRASHES	SHOOT	GUN
WALK	WALKER	TELEPHONE	CALL
DEMAND	DEMANDED	BROOM	SWEEP
HANDLE	HANDLER	WAITRESS	SERVE
CREATE	CREATES	DRIVE	CAR
RELAX	RELAXED	CHAIR	SIT
LOVE	LOVER	LOCK	KEY
SLEEP	SLEEPER	SCISSORS	CUT
NEED	NEEDED	ICE CREAM	LICK
RIDE	RIDER	SHOVEL	DIG
JUMP	JUMPING	LAWNMOWER	MOW
CREDIT	CREDITOR	STRUM	GUITAR

REFERENCES

Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language, 26*, 658-672.

Dannenbring, G. L., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. *Canadian Journal of Psychology, 36*, 435-444.

Feldman L. B. (1987). Phonological and morphological analysis by skilled readers of Serbo-Croatian. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 197-209). London: Academic Press.

Feldman, L. B., & Moskovljević, J. (1987). Repetition priming is not purely episodic in origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 573-581.

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112*, 309-346.

Fowler, C. A., Napps, S. E., & Feldman, L. B. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13*, 241-255.

Gates-MacGinitie Reading Tests (2nd Edition) (1978). Boston: Houghton Mifflin.

Hanson, V. L., & Bellugi, U. (1982). On the role of sign order and morphological structure in memory for American Sign Language sentences. *Journal of Verbal Learning and Verbal Behavior, 21*, 621-633.

Hanson, V. L., & Wilkenfeld, D. (1985). Morphophonology and lexical organization in deaf readers. *Language and Speech, 28*, 269-280.

Henderson, L., Wallis, J., & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma & D. H. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 211-226). London: Erlbaum, 1984.

Henik, A., Friedrich, F. J., & Kellogg, W. (1983). The dependence of semantic relatedness effects upon prime processing. *Memory & Cognition, 11*, 366-373.

Kirsner, K., Brown, H. L., Abrol, S., Chadha, N. N., & Sharma, N. K. (1980). Bilingualism and lexical representation. *Quarterly Journal of Experimental Psychology, 32*, 585-594.

Kirsner, K., Dunn, J., & Standen, P. (1987). Record-based word recognition. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 147-167). Hove and London, UK: Erlbaum.

Kirsner, K., Smith, M. C., Lockhart, R. S., King, M. L., & Jain, M. (1984). The bilingual lexicon: Language-specific units in an integrated network. *Journal of Verbal Learning and Verbal Behavior, 23*, 519-539.

Kroll, J. F., & Borning, L. (1987). *Shifting language representations in novice bilinguals: Evidence from sentence priming*. Paper presented at the meeting of the Psychonomic Society, Seattle, WA.

Lukatela, G., Gligorijević, B., Kostić, & Turvey, M. T. (1980). Representation of inflected nouns in the internal lexicon. *Memory & Cognition, 8*, 415-423.

Meyer, D. E., Schvaneveldt, R, W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 98-118). London: Academic Press.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology, 102*, 963-968.

Napps, S. E., & Fowler, C. A. (1987). Formal relationships among words and the organization of the mental lexicon. *Journal of Psycholinguistic Research, 16*, 257-272.

Padden, C. A. (1987). Grammatical theory and signed languages. In F. Newmeyer (Ed.), *Linguistics: The Cambridge Survey* (pp. 250-266). Cambridge: Cambridge University Press.

Poizner, H., Klima, E. S., & Bellugi, U. (1987). *What the hands reveal about the brain*. Cambridge, MA: Bradford Books/MIT Press.

Poizner, H., Bellugi, U., & Tweney, R. D. (1981). Processing of formational, semantic, and iconic information in American Sign Language. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1146-1159.

Poizner, H., Newkirk, D., Bellugi, U., & Klima, E. S. (1981). Representation of inflected signs from American Sign Language in short-term memory. *Memory & Cognition, 9*, 121-131.

Ratcliff, R., Hockley, W., & McKoon, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. *Journal of Experimental Psychology: General, 114*, 435-450.

Rugg, M. D. (1987). Dissociation of semantic priming, word and non-word repetition effects by event-related potentials. *Quarterly Journal of Experimental Psychology, 39*, 123-148.

Scarborough, D. L., Gerard, L., & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. *Journal of Verbal Learning and Verbal Behavior, 23*, 84-99.

Shand, M. A. (1982). Sign-based short-term memory coding of American Sign Language signs and printed English words by congenitally deaf signers. *Cognitive Psychology, 14*, 1-12.

Shand, M. A., & Klima, E. S. (1981). Nonauditory suffix effects in congenitally deaf signers of American Sign Language. *Journal of Experimental Psychology: Human Learning and Memory, 12*, 464-474.

Smith, C. R. (1975). Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research, 18*, 759-811.

Smith, M. C., Theodor, L., & Franklin, P. E. (1983). The relationship between contextual facilitation and depth of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 697-712.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior, 18*, 399-412.

Stokoe, W. C., Jr., Casterline, D., & Croneberg, C. (1965). *A dictionary of American Sign Language.* Washington, D.C.: Gallaudet College Press.

Supalla, T., & Newport, E. (1978). How many seats in a chair? The derivation of nouns and verbs in American Sign Language. In P. Siple (Ed.), *Understanding language through sign language research* (pp. 91-132). New York: Academic Press.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*, 638-647.

Watkins, M. J., & Peynircioğlu, Z. F. (1983). On the nature of word recall: Evidence for linguistic specificity. *Journal of Verbal Learning and Verbal Behavior, 22*, 385-394.

FOOTNOTES

[*] *Memory & Cognition, 17*, (1989).

[†] Also, IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, New York.

[1] In the case of two stimuli, SAIL and LICK, the ASL signs are usually translated as GO-BY-BOAT and LICK-ICE-CREAM (Supalla & Newport, 1978). Given the consistency of the subjects and items analyses in Experiments 1 and 2, these deviations from the exact translations do not appear to have caused problems.

[2] The term *phonological* need not be limited to use with spoken languages. In the case of American Sign Language (ASL), for example, the term phonology has been used to describe the linguistic primitives related to the visible gestures articulated by the hands, face, and body of the signer.

[3] Experiments 1 and 2 were run in reverse order. The changes in the English + ASL set were necessitated in order to have approximately equal numbers of one- and two-handed sign responses in Experiment 1.

[4] The morphological prime KEY was incorrectly typed as KING in the stimulus list shown in Experiment 2 only. The target word LOCK was therefore omitted from analysis in this experiment.

Anomalous Bimanual Coordination Among Dyslexic Boys[*]

Marshall Gladstone,[†] Catherine T. Best,[††] and Richard J. Davidson[†††]

Most neuropsychological models of dyslexia have focused on left hemisphere dysfunctions; however, several recent studies also suggest difficulties with interhemispheric integration. We assessed that possibility by testing bimanual coordination in 18 dyslexic and 18 nondisabled dextral boys (9-14 years) of above average intelligence on an Etch-a-Sketch-like apparatus. The task measured various ratios and directions of hand rotation with and without visual feedback. Group performance was equivalent on trials requiring parallel movements of the hands (both clockwise). In contrast, on trials requiring mirror movements (clockwise + counterclockwise), dyslexics were slower and less accurate than controls, particularly with their left hands. Moreover, dyslexics had notable difficulty on the mirror movement trials without visual feedback, often unknowingly reverting to parallel hand movements. The dyslexics' deficits were largely consistent with previous observations of partial commissurotomy patients, supporting the hypothesis of impaired interhemispheric coordination. However, the inability to perform mirror movements seen among the dyslexics has never before been observed in any other population. We propose that a combination of deficient interhemispheric collaboration plus anomalous organization of the ipsilateral motor hand pathways may account for the dyslexics' performance.

Recent theoretical proposals suggest that efficient interhemispheric integration is important in reading development (Gladstone & Best, 1985; Kershner, 1985a), and that dyslexic children may have deficits in interhemispheric collaboration (Davidson, Leslie, & Saron, 1988; Denckla, 1986; Gladstone & Best, 1985; Kershner, 1985b). A number of empirical investigations lend support to this proposal. Compared to normal readers, reading impaired children show less interhemispheric EEG coherence (Sklar, Hanley, & Simmons, 1973). Anatomical measures of the corpus callosum, the major neocortical commissure, have been reported for two dyslexic brains; both showed abnormal thinning (Drake, 1968; Witelson, cited in McGuinness, 1985). Behaviorally,

Haskins Laboratories **SR-97/98**

Status Report on Speech Research *1989*

certain tachistoscopic (Gross-Glenn & Rothberg, 1984; Yeni-Komshian, Isenberg, & Goldberg, 1974) and dichotic (Kershner, Henninger, & Cooke, 1984; Obrzut, Obrzut, Hynd, & Pirozzolo, 1981) studies of reading-disabled children have also suggested abnormal interhemispheric sharing of sensory information. Whereas normal age-matched children display a stable right-ear advantage on verbal dichotic tests, reading-disabled children switch the direction of ear advantage as a function of which ear is attended (Obrzut et al., 1981) and whether the response mode is oral or written (Kershner et al., 1984). This "attentional switching," a phenomenon usually seen only among younger normal children, has been attributed to the inadequate regulation of attentional resources between the hemispheres (Hiscock & Kinsbourne, 1980).

Thus, while contemporary views of dyslexia have more commonly attributed reading failure to left hemisphere impairment (Rudel, 1985), recent evidence has also implicated deficient interhemispheric attentional regulation and sensory integration. However, a number of questions about interhemispheric coordination among disabled readers remain unanswered. For example, relatively little is known about the role of interhemispheric processes in the bilateral motor anomalies that have been clinically observed for many years in reading-disordered children (Denckla, 1985; Tomaino, in press). Recently, Denckla (1986) has stated that "Orton's concept of mixed dominance may be important not in terms of 'competing engrams' in the visual field, but rather in terms of [interhemispherically] competing motor coordination systems in the anterior portions of the brain." (p. 193). In light of the theory and data suggesting impaired interhemispheric communication among dyslexics, it would be important to determine whether their atypical motor performance is also indicative of anomalous interhemispheric processes.

The motor anomalies of "pure dyslexic" children (i.e., children with primary reading impairment without other significant language, neurodevelopmental or emotional disorders) appear to be largely confined to fine and/or repetitive movements of the distal musculature of the fingers and hands. For example, impairment in rapid alternating finger movements was the only distinguishing motor characteristic found in one group of dyslexics (Owen, Adams, Forrest, Stolz, & Fisher, 1971). Similarly, dyslexic children have shown difficulties in a sequential finger to thumb opposition task (Denckla, 1985). Of particular importance to the issue of interhemispheric collaboration are findings that indicate poor coordination between the hands. Notably, certain types of motor imprecision among dyslexics appear only when bilateral coordination is required (Badian & Wolff, 1981; Denckla, 1985; Wolff, Cohen, & Drake, 1984). When unilateral movement is required, synkinetic or mirror movements may occur in the opposite hand (Denckla, 1985). Moreover, there is often a large left-right asymmetry, with poorer than normal left hand performance among right-handed dyslexics (Leslie, Davidson, & Batey, 1985).

Any of these motoric findings may implicate disturbances in interhemispheric coordination or left hemisphere processes. Manual performance in dextral brain-damaged populations has suggested that the left hemisphere is dominant for fine motor control of *both* hands, influencing the left hand via callosal connections to the right hemisphere motor centers (Geschwind, 1975). Thus, deficits in left hand performance have been found in both left hemisphere-damaged individuals (Wyke, 1971a) and commissurotomy patients (Kreuter, Kinsbourne, & Trevarthen, 1972). The impaired left hand performance of dextral dyslexics suggests the possibility of poor collaboration between the hemispheres (Denckla, 1986; Leslie et al., 1985) or impaired left hemisphere function (Rudel, 1985).

When the motor deficits of reading-disabled children are exclusive to conditions requiring bimanual coordination, interhemispheric processes are implicated. Specifically, deficits on an alternating bimanual tapping task have been interpreted as evidence of impaired interhemispheric coordination (Badian & Wolff, 1977; Klicpera, Wolff, & Drake, 1981). The general relation between intermanual coordination and interhemispheric collaboration has been most clearly demonstrated by Preilowski (1972, 1975). In this study, epileptic adults whose anterior neocommissures had been sectioned were compared to normal and unoperated epileptic adults. The subjects were required to learn a bimanual coordination task similar to Etch-a-Sketch. While seated in front of a screen, participants were required to turn two cranks in order to guide a stylus along a prescribed pathway. Control subjects learned to coordinate their hands to produce a relatively straight line within the required diagonal pathway. In contrast, the commissurotomy patients failed to develop the smooth intermanual motor control seen among other subjects. Rather, they appeared to rely on visual feedback mechanisms to correct their movement errors. Moreover, when visual feedback was removed, the control subjects were able to retain their learned pattern of intermanual coordination. In contrast, the commissurotomy patients were largely unable to do so, indicating their reliance on inefficient visual feedback. Further, they demonstrated particular difficulty controlling their left hands. Preilowski's findings indicate the specific role of the anterior portion of the corpus callosum in left hand control and the acquisition of intermanual coordination.

Impaired intermanual coordination, however, has also been associated with specific left hemisphere damage. Left hemisphere damaged patients produced more errors on a bimanual pantograph task than did normals or right hemisphere damaged patients (Wyke, 1971b). Although Wyke's task did not assess individual hand contributions to overall bimanual ability, since a pantograph links the two hands to control a single stylus, her data nonetheless suggest a left hemisphere contribution to bimanual ability.

In light of the evidence for both left hemisphere and interhemispheric contributions to manual and intermanual coordination, it is not clear to what extent the manual coordination deficits of dyslexics may derive from left hemispheric and/or interhemispheric anomalies. Wolff and his coworkers have attempted to address this issue (Badian & Wolff, 1977; Klicpera et al., 1981; Wolff et al., 1984). They assessed "reading retarded" boys on a tapping task that required them to maintain various metronome entrained tapping rates. Unimanual, synchronous bimanual, and alternating bimanual modes of this timed tapping precision task were utilized. This procedure permitted the measurement of individual hand contributions to various modes of manual coordination. Specific left hand deficits appeared only in the bimanual alternating mode, which the authors interpreted as evidence of impairment in rapid communication between the hemispheres (Badian & Wolff, 1977; Klicpera et al., 1981). Since the bimanual condition alone revealed deficits, the results seem analogous to Preilowski's (1972) findings with partial commissurotomy patients. However, a later report by Wolff and colleagues (Wolff et al., 1984) failed to support their earlier conclusion. That study found that when the absolute *rate* of motor production was measured on a variety of speeded motor tasks, including some unrelated to bimanual coordination, the performance of the reading-impaired children was better characterized by the rate at which motor timing broke down, rather than by intermanual ability *per se*. It appeared that the rapid motor timing ability of the dyslexics was impaired, and that this had been most evident in the bimanual alternating tapping task. Consequently, their use of a rapid timing precision task confounded intermanual coordination ability and rapid motor timing control. Given the evidence that the left hemisphere is specialized for rapid motor sequencing ability

(Kimura, 1977), the Wolff et al. (1984) report suggests that the earlier observed intermanual anomalies of poor readers may not be due specifically to interhemispheric collaboration deficits. It should be noted that Davidson and his colleagues (Davidson et al., 1988) have proposed that problems in the timing of interhemispheric transfer in dyslexics may actually interfere with left hemisphere response processes and thus reveal themselves as left hemisphere deficits. No study to date has shown the differences in intermanual coordination between normal and impaired readers to be independent of rapid motor sequencing ability. Clarification of this issue might shed some light on the role of interhemispheric processes in developmental dyslexia. Consequently, in the study reported here, the performance of dyslexic and nondyslexic boys was compared on a bimanual fine motor coordination task designed to measure manual abilities independent of absolute timing precision. Following Preilowski (1972), we utilized a task similar to Etch-a-Sketch to assess bimanual performance, but designed our apparatus to measure the independent contributions of left and right hands. Also following Preilowski, both visual feedback and no visual feedback conditions were used to evaluate the role of visual guidance in the performance of this bimanual task.

Method

Subjects

Eighteen impaired (mean age = 12.08 yr.; range = 10.0 to 14.10 yr.) and 18 nonimpaired readers (mean age = 12.31 yr.; range = 9.8 to 14.7 yr.) were selected from schools in Westchester County, New York. All children were right handed boys (as assessed by the Harris Test of Lateral Dominance, Harris, 1958) and participated in a previously published study of dyslexia (Leslie et al.,1985). Each child was paid ten dollars per research session. Children with histories of head trauma, seizures, psychiatric illness, recent use of psychotropic medication or uncorrected visual or auditory acuity deficits were excluded from the study. Subjects were of at least average intelligence (WISC-R > 90; Block Design Subtest > 7). Reading level was measured by word recognition (Woodcock Reading Mastery Test: Woodcock, 1977). Reading impairment was determined by the Myklebust Quotient [(2 x reading age) /(mental age + chronological age)]. Impaired readers had Myklebust quotients below .85, while controls scored above .95. Additionally, all impaired readers were deficient in object naming [Neurosensory Center Comprehensive Examination for Aphasia (NCCEA), Spreen & Benton, 1977] or in Rapid Automatized Naming (Denckla & Rudel, 1976). None of the control subjects had naming deficits.

Table 1 lists the group differences on tests related to the inclusion criteria.

Apparatus

A Hewlett-Packard model 7035B x-y plotter was used to assess bimanual coordination. The plotter pen was controlled by two small, remote 10 turn hand-manipulated potentiometers, with one knob controlling the x axis and the other controlling the y axis. The apparatus functioned much like the children's game Etch-a-Sketch. The position of the potentiometers was counterbalanced so that the left hand controlled the x axis for half of the subjects in each group, but controlled the y axis for the other half. The movement of the two knobs allowed the subject to steer the plotter pen through various predetermined pathways. Additionally, an opaque screen was hinged to the base of the plotter to occlude the subject's view of the pen and plotter during the no-visual feedback trials.

Potentiometer movement was monitored by optical counters mounted below the potentiometers. The counters were triggered by movement of the gear teeth in the knobs, and fed two decade counters (Thornton Model DEC 110). Knob movement of approximately 12° rotation, in either direction, registered as a single count on the LED readout of the decade counter. At the end of a trial, the readout value was recorded by hand.

To assess memory for line orientation, as a control measure for the no visual feedback condition (see Procedure), a simple apparatus was used. It consisted of a 11 in. x 14 in. (280 mm x 350 mm) Masonite board onto which a 8-1/2 in. x 11 in. (243 mm x 280 mm) lucite frame had been glued to hold a sheet of paper of the same size. A clear lucite pointer with a black line scribed in it was attached so as to pivot from the midpoint of the base of the frame.

TABLE 1. Scores for criterion measures and reading tests.

	DYSLEXIC	CONTROL
Wechsler Intelligence for Children-Revised (Verbal IQ)	111.68[a] (12.86)[b]	126.84 (11.69)
(Performance IQ)	119.74 (11.95)	116.79 (13.87)
(Full Scale IQ)	117.21 (12.50)	124.84 (12.28)
Woodcock Reading Mastery Word Identification %	17.89 (16.94)	81.79 (10.79)
Harris Laterality %	96.11 (8.50)	97.22 (4.61)
Rapid Automatized Naming: Colors %	43.94 (35.88)	80.39 (15.19)
Rapid Automatized Naming: Objects %	33.11 (23.60)	72.50 (14.04)
Rapid Automatized Naming: Numbers %	18.44 (26.44)	77.50 (21.54)
Rapid Automatized Naming: Letters %	10.11 (21.22)	82.78 (16.29)
NCCEA Visual Naming	61.78 (29.53)	75.29 (11.14)
GFW Auditory Disc.: quiet %	92.00 (16.26)	73.39 (24.40)
Myklebust Reading Quotient	72.00 (.09)	1.21 (.17)

[a] IQ test expressed in standard score units (M=100, SD=15); all other tests except Myklebust quotient expressed as percentiles.
[b] SD in parentheses.

Stimuli

Each stimulus for the Etch-a-Sketch-like task consisted of a "path," defined by parallel straight lines drawn 5 millimeters apart on 8 1/2 in. x 11 in. (243 mm x 280 mm) white paper, which was positioned on the plotter screen. The subject was to guide the plotter pen between the two lines. There were two test conditions: Visual Feedback (VF) and No Visual Feedback (NVF).

In the VF Condition the paths were 180 mm in length, varying only in the angle of line orientation on the page. Six trials were presented: a 26.5° angle and its supplementary and complementary angles (63.5°, 116.5° and 153.5°), as well as a 45° angle and its supplement (135°). The 26.5° angle required a bimanual ratio of 2:1 for left to right hand (or right to left hand) movement of the potentiometer knobs. At 26.5°, both hands moved in a clockwise direction. At 63.5°, the hand ratio for each subject was reversed (1:2), with both hands moving in a clockwise direction. At 116.5° and 153.5°, the left to right hand ratios were also 1:2 and 2:1, but one hand turned clockwise while the other turned counterclockwise. At 45°, the two hands turned clockwise at the same rate (1:1 ratio). At 135° the 1:1 ratio was retained, but the hands turned in opposite directions (clockwise and counterclockwise). The six angles thus assessed equal and unequal rates of left and right hand movement, as well as unidirectional (clockwise) and bidirectional (clockwise/counterclockwise) movements. In the analyses, the former factor is designated as *movement* ratio, and the latter as *direction*.

The stimuli for the NVF Condition were identical to those for the VF Condition, except that a line had been drawn perpendicular to the pathway, 60 mm from the start of the 180 mm pathway, in order to mark the beginning of the no-visual feedback portion of the trial.

Procedure

The two bimanual tasks thus comprised a total of 12 trials presented to each subject. The two conditions were presented to all subjects in the same order, VF followed by NVF. Both conditions were preceded by a single practice trial. Within both the VF and NVF conditions, presentation order of pathway orientations was randomized across subjects. At the conclusion of all trials, the memory of line orientation task was administered.

At the beginning of the session, the subject was seated in front of the plotter screen and situated so that each potentiometer knob could be controlled with the thumb and index finger of the assigned hand. All subjects were allowed approximately one minute of practice with the apparatus to familiarize themselves with the movement of the plotter pen. Prior to the start of each trial the counters were set to zero. During each bimanual trial, verbal encouragement was given freely, but no assistance with the task nor feedback about performance was provided until the trial ended. Subjects were instructed to steer the plotter pen through different pathways quickly, with an emphasis on accuracy in staying between the lines. They were told that if they strayed outside the lines, they should try to reenter the pathway at the point where they had exited it. Time and individual left and right hand scores were recorded at the conclusion of each trial.

VF Condition

For this condition, the subjects steered the plotter pen through the six straight pathways in free viewing. At the beginning of each trial, the examiner placed the sheet containing one of the blank pathways on the plotter screen and moved the pen to the beginning of the pathway. Time was recorded by stopwatch to the nearest .1 sec, beginning from when the subject began steering the pen. Left and right hand potentiometer readings were recorded from the decade counters for each trial.

NVF Condition

At the conclusion of the VF trials, subjects were given a brief rest and were then given the six NVF trials, preceded by a practice trial. They were instructed to begin as in the VF trials, but to stop moving the pen momentarily when it reached the small perpendicular line drawn across the path, so that the experimenter could lower the screen that occluded the subject's subsequent view of the plotter. This screen was marked with an "X" in the center as a visual fixation point, approximately at eye level, to keep the subjects from watching their hands as they completed the path without visual feedback. Time and potentiometer output scores were recorded separately for the first 60 mm (visual feedback) and last 120 mm (no visual feedback) portions of each trial. When the subjects reached either the end of the path or a point 120 mm from the start of the no visual feedback portion of the trial (determined by an arc drawn 120 mm from the origin of that segment of the path), the subjects were stopped and the time and potentiometer readings were recorded. Subjects were permitted to see their performance at the conclusion of each trial, but no cues about performance were given during the trial.

Memory for Line Orientation

To assess the effect of visual memory for line orientation on performance in the NVF condition, a separate line orientation memory task was administered at the conclusion of the session. Subjects were presented with a sheet of paper placed in the apparatus described earlier, onto which was drawn a line segment starting from the midline of the bottom of the page and oriented at one of the six stimulus angles. Subjects were instructed to remember the orientation of the line. Following a 15 sec presentation and a 10 sec delay, subjects were represented with a blank sheet of paper in the same apparatus. They were asked to reproduce the orientation of the line segment by pivoting the lucite pointer to the correct angle. The subjects' response in angular degrees was recorded for each of the six trials. Presentation order for the six line segments was randomized across subjects.

RESULTS

VF Condition

Error Scores

For each trial, in each condition, an ideal score (mm) was computed to reflect the minimum x and y axis inputs necessary to complete the path without deviation from the center. The subject's obtained scores were compared to these ideal scores on each trial: error scores for each hand were calculated as obtained score minus ideal score for the x and the y axis.

A 2 X 2 X 3 X 2 analysis of variance (ANOVA) was computed for error scores in the VF Condition with Group as a between-subjects factor and Hand, Movement Ratio, and Direction as within-subjects factors. "Direction" refers to the direction of bimanual movement required for the trial, (unidirectional vs. bidirectional) and "Movement Ratio" refers to the ratio of right to left hand movement required (1:1, 1:2, or 2:1) (see **Stimuli**).

There were no main effects for Group, Movement Ratio, Hand, or Direction. However, the interaction of Group X Direction, $F(1,34) = 4.88$, $p < .04$, and Group by Direction by Hand, $F(1,34) = 4.1$, $p = .05$, were both significant. Figure 1 suggests that the Group by

Direction interaction was attributable to a greater error rate by dyslexics in the bidirectional trials, and greater error rate by controls in the unidirectional trials. However, the simple effects tests for this two-way interaction were nonsignificant.

In contrast, the simple effects tests were significant for the Group X Direction X Hand interaction, which was attributable to Group X Direction differences in left hand performance. The control subjects showed significantly greater left hand errors in the unidirectional trials than in the bidirectional trials, $F(1,34) = 5.99$, $p < .02$. The reverse effect for the dyslexics (see Figure 1) was not significant, but they did show significantly greater errors for the left hand than the right hand in the bidirectional trials, $F(1,34) = 5.15$, $p < .03$. However, there was no group difference in left hand performance on either the bidirectional or the unidirectional trials.

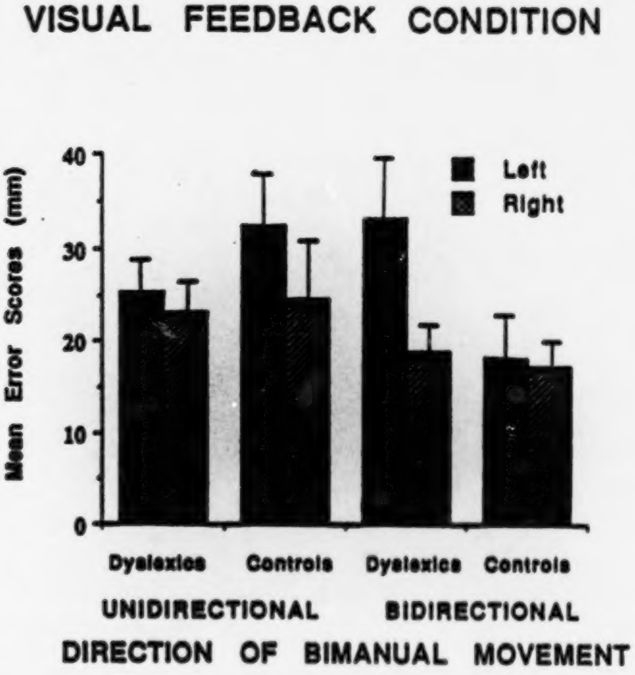


Figure 1. Mean error scores in the Visual Feedback (VF) condition. Error bars represent standard error of the mean.

Completion Time Scores

The analysis of completion time also indicated group differences (see Figure 2). A 2 X 3 X 2 ANOVA, with Group, Movement Ratio, and Direction as factors, failed to find a significant main effect for Group. There was a significant effect for Direction, $F(1,34) = 5.14$, $p < .03$, and for the interaction of Group X Direction, $F(1,34) = 5.09$, $p < .03$. Although Movement Ratio was also significant, $F(2,68) = 12.69$, $p < .001$, with longer completion times associated with equal rather than unequal hand movement ratios, this did not interact with either Group or Direction X Group.

Simple effects tests of the Group X Direction interaction indicated that dyslexics were slower than controls in the bidirectional trials only, $F(1,34) = 4.07$, $p = .05$. While the dyslexics showed no completion time differences between Directions, the controls were faster in the bidirectional than the unidirectional trials, $F(1,34) = 10.23$, $p = .003$. Thus, in the visual feedback trials, the dyslexics were slower than controls only on trials requiring opposing hand movements.

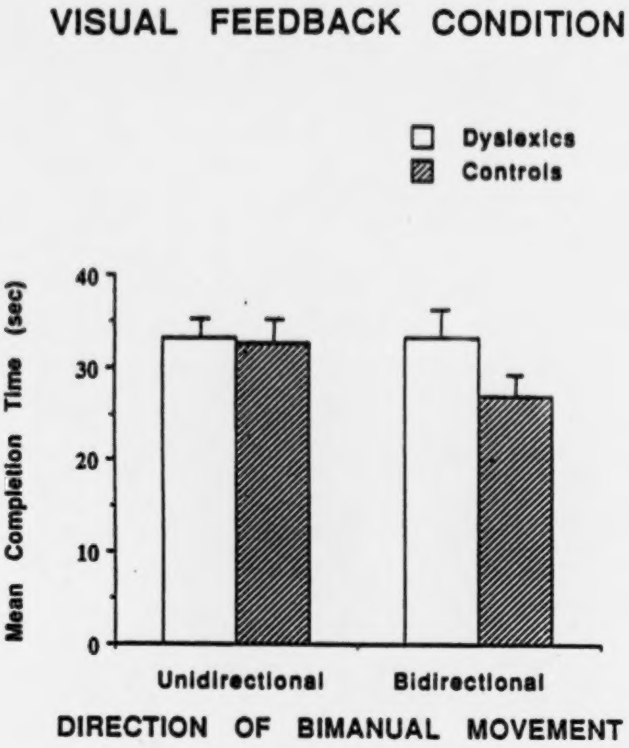


Figure 2. Mean completion times for the Visual Feedback (VF) condition.

NVF Condition

Error Scores

In the NVF trials, the last 120 mm of the pathway was performed without visual feedback, and therefore relied primarily on memory of performance during the first 60mm. Subjects' error scores on this last 120mm were analyzed by a 2 X 2 X 3 X 2 ANOVA, with Group as a between-subject factor, and Hand, Movement Ratio, and Direction as within-subject factors, as in the VF condition. There were no main effects for Group or Direction, but Movement Ratio, $F(2,68) = 14.96$, $p < .001$, and Hand, $F(1,34) = 5.09$, $p < .04$, were both significant. Overall, the right hand produced fewer errors than the left, and trials requiring a 1:1 ratio of input from the two hands (45° and 135°) produced fewer errors than 1:2 or 2:1 ratios. The latter effects did not, however, significantly interact with Group, and thus were not specific to the dyslexics.

Although this analysis failed to indicate any significant Group differences, we noticed that some subjects completely reversed the direction of rotation of one or both hands after visual feedback was removed, and that this pattern appeared strikingly different for the dyslexic versus the control subjects (see Figure 3). Since the hand error scores were computed from the amount of hand input to the potentiometers, regardless of direction, such reversals could yield a near zero error score by a turn of the potentiometer knob of the correct magnitude, but in completely the wrong direction. These directional reversals can be treated in two different ways, which led us to develop two new measures of error. First, we computed the arc degrees of difference between the subject's response line and the center of the prescribed path (see Figure 3), referred to as *arc error* scores. Second, we noted that directional reversals placed the response line in the spatial quadrant opposite to the stimulus path. For example, if the stimulus path was at 135°, which is in the negative x, positive y quadrant, and the response line was in the positive x, positive y quadrant, the subject had reversed the direction of rotation of the hand controlling the x axis. Therefore, reversal frequency served as another dependent measure.

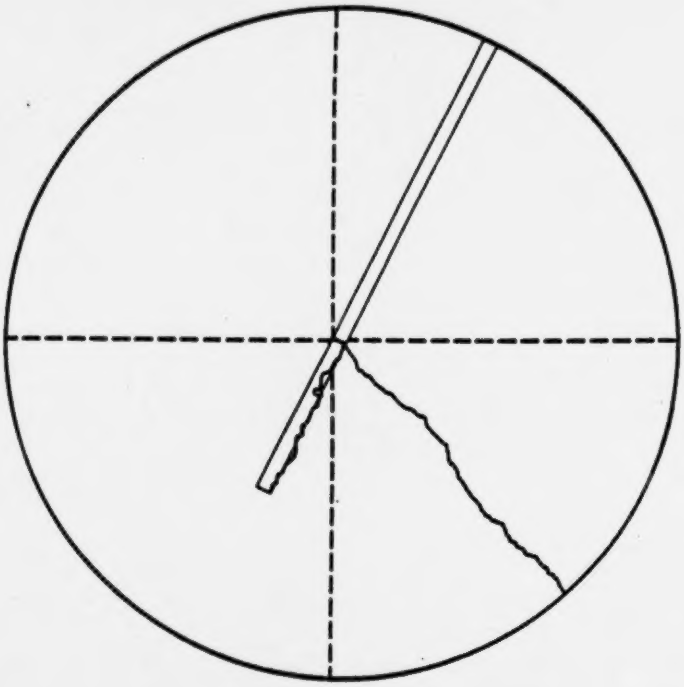


Figure 3. Tracing of a reversal error produced by a dyslexic subject on one of the bidirectional NVF trials.

Arc Error Scores

Arc error scores were analyzed in a 2 X 3 X 2 ANOVA, with Group as a between-subject factor, and Movement Ratio and Direction as within-subject factors. There was a significant main effect for Group, $F(1,34) = 14.61$, $p < .005$, but not for Movement Ratio or Direction. The interaction of Group X Direction was significant, $F(1,34) = 8.96$, $p = .005$) Simple effects tests on this interaction revealed that the dyslexics had a greater error magnitude on the bidirectional trials in comparison to the controls, $F(1,34) = 15.09$, $p < .0005$, and in comparison to their own unidirectional performance, $F(1,34) = 12.17$, $p < .002$ (see Figure 4).

Reversal Frequency Scores

Reversal frequency scores were submitted to binomial probability tests, which yielded no group differences for any of the unidirectional trials. However, for each of the bidirectional trials, dyslexics produced significantly more reversals than controls (see Table 2).

TABLE 2. Number of reversals in bidirectional trials of the No Visual Feedback (NVF) Condition.

Trial Angle	Dyslexics	Controls	Binomial Probability
116.5°	7	0	.008
135.°	8	0	.004
153.5°	5	0	.031

This finding, in conjunction with the nonsignificant Group effect in the hand error scores, indicates that the group differences in arc error were largely a function of reversal frequency on bidirectional trials. Group differences attributable to this combination of large arc and low hand error scores are possible only when the errors are caused by complete reversals in the direction of hand rotation.

To assess whether the reversals were preferentially influenced by one hand or by one direction of movement (clockwise vs. counterclockwise), we examined the quadrant of actual response on each trial for which a reversal occurred (Figure 4). Among the 24 reversals found in the dyslexic group (20 in the bidirectional trials, 4 in the unidirectional trials), 13 were attributable to the left hand alone (11 on bidirectional trials, 8 of these involving a switch from counterclockwise to clockwise motion), 10 to the right hand (8 on bidirectional trials, 5 of these involving a counterclockwise to clockwise switch), and 1 to both hands (bidirectional trial). According to binomial tests, neither the hand difference nor the difference in direction of switching was significant. Nor was there a significant association between hand and switching direction, by a Fischer's Exact Probabilities test. However, the dyslexics *were* more likely to revert to unidirectional (parallel) movement of the two hands (19 instances) than to revert to bidirectional (mirror) movements (5 instances), $p < .003$ by binomial test. Moreover, reversals to unidirectional (parallel) movement were more likely to result in clockwise motion of both hands (13 instances) than in counterclockwise motion (6 instances), $p = .05$ by binomial test. Only three reversal errors occurred among the control subjects, all in unidirectional trials, that is, reverting to mirror movement.

It is notable that 15 of the 18 dyslexic subjects produced reversals at least once, whereas only two of the 18 controls did so. Seven of the dyslexics, and one control, reversed on more than one trial, even though they had been shown their errors at the end of each trial. Thus, awareness of their reversal errors did not stop a number of dyslexics from reversing on subsequent trials.

Completion Time Scores

Trial completion times were also analyzed, by a 2 X 3 X 2 ANOVA with Group as a between-subject factor and Movement Ratio and Direction as within-subject factors. Results indicated no significant main effects. Although the dyslexics appear to have performed somewhat more slowly than the controls on the bidirectional trials (36.62 sec. vs. 28.54 sec.) and on the unidirectional trials (33.02 sec. vs. 29.7 sec.), the effect was not significant. However, the interaction of Group X Movement Ratio was significant, $F(2,62) = 3.72$, $p < .04$. This effect was the result of the dyslexics' slower performance on the 45° (unidirectional, 1:1 ratio of hand inputs), 135° (bidirectional, 1:1 ratio), and 153.5° (bidirectional, 2:1 ratio) trials. No other completion time effects were significant.

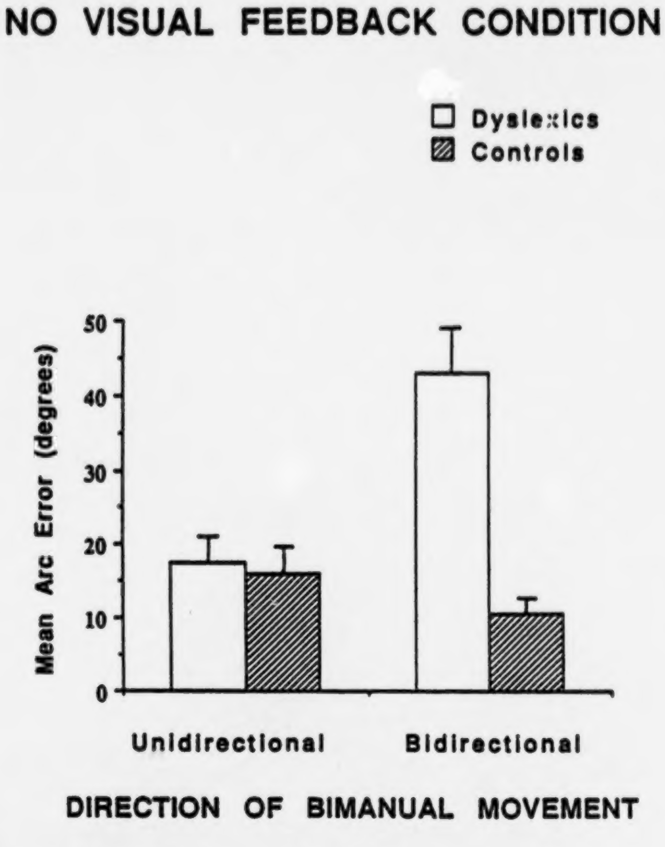


Figure 4. Mean arc error scores for the No Visual Feedback (NVF) condition.

Several findings in the NVF condition thus parallel those found in the VF Condition. Dyslexics were slower and less accurate on bidirectional trials, and showed a bias toward greater left hand errors. However, in contrast to the VF condition, hand error scores did not reveal significant Group X Direction differences for the NVF condition. The *lack* of group differences in NVF hand errors can be accounted for by the incidence of reversals among dyslexics, which *were* evident in arc error scores and reversal frequencies.

Counterbalancing of Hand and Plotter Axis

The effect of counterbalancing hand of control with plotter axis was examined by analysis of variance for the VF and NVF trials. There were no significant effects in any condition.

Memory for Line Orientation

For the line memory task, error scores were computed as the average difference in angular degrees between the target lines and the subjects' responses on the six trials. Scores among the dyslexics appeared to contain several extreme values, so group scores were tested for homogeneity of variance. Since the group variances were significantly different, $F(1,34) = 5.62$, $p = .023$, a t-test for unpooled variances was conducted to compare group means. There was no significant group difference in memory for line orientation.

Age as a Factor

In order to assess the influence of age on group performance, age at time of testing was entered as covariate with the dependent measures of time and left and right hand error scores in the VF condition, and arc error scores in the NVF condition. The ANOVA indicated no significant effect for age in any condition.

DISCUSSION

In the present study, the performance of dyslexic children on an intermanual coordination task unrelated to the reading process was consistent with the well-accepted position that they have atypical functional cerebral organization. The disabled readers showed impairments in bimanual coordination, relative to their nondisabled controls, under several conditions. Group differences were evident on both speed and accuracy measures; impaired readers performed more slowly and less accurately with visual feedback, particularly with their left hands during bidirectional trials, and also performed less accurately without visual feedback. A primary difference between the two conditions was that under visual feedback monitoring, the dyslexics were able to correct the direction of their left hand movement. Their high error scores, however, reflect the inefficiency of this approach. In contrast, without visual feedback, these subjects were unable to monitor their hand movements visually, and unknowingly reverted to unidirectional bimanual movement. This was demonstrated in both their arc error and frequency of reversal scores.

Most striking was the tendency of nearly all the dyslexics, but none of the control subjects, to revert unknowingly to unidirectional (parallel movements) of the two hands under requirements of bidirectional (mirror) movements without visual feedback. This pattern of reversals more often resulted in clockwise, rather than counterclockwise, movements of both hands. Thus, the dyslexics had difficulty maintaining mirror-image movements, and/or showed a strong preference for unidirectional bimanual movements, usually clockwise.

At a general level, the impaired intermanual coordination of our dyslexic subjects coincides with Preilowski's (1972,1975) report of intermanual deficits among partial commissurotomy patients. Both the commissurotomy patients and our disabled readers showed greater left- than right-hand impairment, which Preilowski attributed to a failure in the interhemispheric coordination of motor commands from the dominant left hemisphere in his subjects. In addition, both these groups had difficulty maintaining certain preestablished bimanual movements without visual guidance, and unknowingly reverted to movement patterns that were presumably less demanding for them. Preilowski explained this behavior in his commissurotomy patients as a failure to develop efficient motor feedforward, a putative function of the anterior portion of the corpus callosum. The parallels in the two sets of findings suggest that there may be similar interhemispheric deficiencies among dyslexics.

The present results are also consistent with other reports of impairments among poor readers on intermanual (Badian & Wolff, 1977; Klicpera et al., 1981) and left hand (Leslie et al., 1985) performance, which were interpreted as evidence of deficient interhemispheric processes. However, as we noted in the Introduction, Wolff et al. (1984) found that impaired readers differed from normal controls in timing precision on a variety of speeded motor tasks, and not only on bimanual coordination. The tapping task used in the earlier Wolff studies had thus confounded intermanual coordination with timing precision, which is dependent on specialized left hemisphere processes (Kimura, 1977; Vaughn & Costa, 1962; Wyke, 1971a). These considerations suggest the possibility that the poor readers' bimanual tapping deficit reflected a left hemisphere rather than an interhemispheric dysfunction, thus limiting comparisons between Wolff's and Preilowski's findings.

In contrast, the present study assessed intermanual coordination independent of absolute timing precision. Therefore, the bimanual impairments we found in reading-disabled boys cannot be ascribed solely to deficits in left hemisphere control of rapid sequential motor output. However, it is premature to draw the alternative conclusion that interhemispheric processes alone are responsible.

A primary limitation in interpreting our findings is that there are no previous reports of impaired performance on bidirectional, or mirror-like, movements in any population. Thus, there is no precedent for relating the bidirectional movement deficits to hemispheric or interhemispheric function. No such deficits are present in Preilowski's studies with commissurotomy patients; in fact, both his patients and normal control subjects showed *better* performance on bidirectional, mirror movements than on unidirectional, parallel ones (see also Sperry, Gazzaniga, & Bogen, 1969). However, it should be noted that Preilowski's task required movements of the lower arms (proximal musculature) to turn cranks while ours required fine movements of the thumb and forefinger (distal musculature) to turn small knobs. Therefore, differences in the pattern and degree of cortical control over the distal versus proximal musculature (Brinkman & Kuypers, 1972) might seem to suggest an explanation for the disparity between our findings and those of Preilowski.

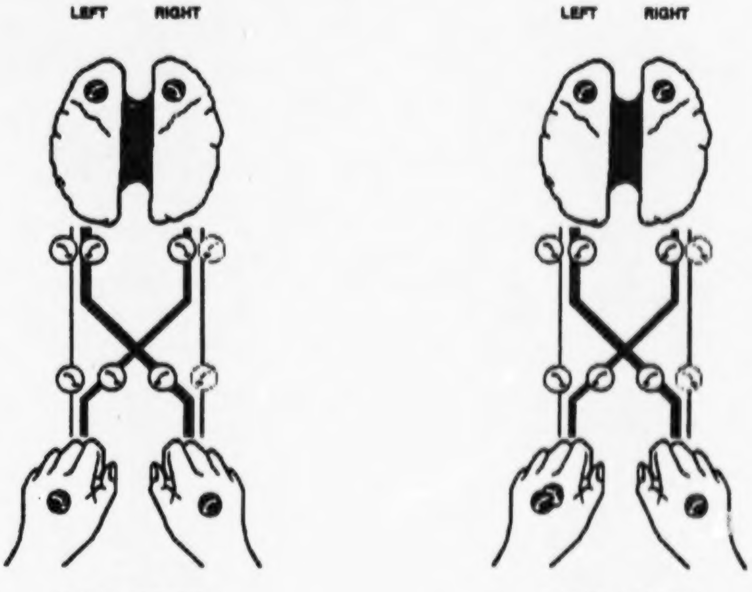
Two observations, however, mitigate against the possibility that our findings completely depend on the requirement for distal motor control. First, our control subjects, like Preilowski's control and commissurotomy subjects, demonstrated clear preferences for mirror movements. They were faster and more accurate on bidirectional (mirror movement) than unidirectional (parallel movement) trials with visual feedback, and the few reversals they produced without visual feedback all occurred on unidirectional trials (parallel movement requirement) and resulted in mirror movement of the hands. Second, using unimanual finger-thumb apposition tasks, Denckla (1985) found overflow of mirror movements in the contralateral hand among "pure" dyslexics. The difference between our finding and Denckla's needs explanation, given that both involved distal motor control. In this regard, it should be noted that her study employed a unimanual rather than bimanual coordination task to examine mirror movement preferences. Based on these two considerations, our findings with dyslexics appear to be specific to the *relationship* between their distal motor coordination abilities and our bimanual task requirements.

Can the anomalous parallel bimanual movement tendencies observed among our dyslexics be explained by existing neural models of the more commonly observed mirror movement patterns? Current theory argues that both commissural action and the distribution of ipsilateral (uncrossed) vs. contralateral (crossed) pathways from motor cortex in each hemisphere to each hand contribute to the appearance or inhibition of mirror movement (Haerer & Currier, 1966; Nass, 1985; Preilowski, 1975). Evidence from unilateral brain-damaged and hemispherectomy patients suggests that the ipsilateral motor pathway from each hemisphere contains a mirror representation of the movement pattern projected via the larger contralateral pathway. This would lead to an underlying preference for mirror movements, which is normally suppressed because the ipsilateral message is overridden by the contralateral message transmitted via the corpus callosum from the opposite hemisphere. Increased mirror movement tendencies in congenital acallosals (Dennis, 1976), and in commissurotomized patients (Preilowski, 1975), indicate that corpus callosum plays an important role in this suppression of mirror movements.

Figure 5a illustrates the neural model proposed by Preilowski (1975) to account for the mirror movement preference in normal subjects. In our diagram, the solid black line descending from the left hemisphere to the right hand (crossed or contralateral pathway) and to the left hand (uncrossed or ipsilateral pathway) represents the dominant influence of the left hemisphere on bimanual performance. The dark gray bridge between the hemispheres represents the corpus callosum; the arrow within the corpus callosum indicates that the left hemisphere has a dominant influence over the right hemisphere for manual control via the corpus callosum. The greater thickness of the contralateral than of the ipsilateral pathways illustrates the greater degree of contralateral influence over manual control. The pathways emanating from the left hemisphere are darker and thicker than those emanating from the right hemisphere, to illustrate left hemisphere dominance in manual control. The arrows within the circles represent commands for the direction of hand rotation. Note that the ipsilateral pathway from each hemisphere is assumed to represent the mirror image motion of that specified by the contralateral pathway from the same hemisphere, following Preilowski (1975).

Figure 5a thus illustrates that for normal subjects, bidirectional (mirror) movements of the hands should be relatively easy because each hand receives identical movement commands through its contralateral and ipsilateral pathways. However, unidirectional (parallel) movements should be more difficult due to the conflicting commands to each hand through the contralateral versus the ipsilateral pathways. In the latter case, there is greater conflict for the left hand than for the right hand because there is greater ipsilateral influence from the dominant left hemisphere to the left hand than there is from the nondominant right hemisphere to the right hand.

This model of mirror movement tendencies, exemplified in Preilowski (1975) and Nass (1985), characterizes quite well the performance of various clinical populations as well as normal children and adults under certain task demands. It cannot, however, account for our dyslexic subjects' reversals to unidirectional, parallel action of the hands without some modification. A general hypothesis of poor interhemispheric coordination would be consistent with the unwanted appearance of an underlying movement preference under increased task demands, such as removal of visual feedback, which would indicate deficient inhibition of the ipsilateral messages. Moreover, it is in line with other suggestions of abnormal interhemispheric functions in disabled readers (e.g., Davidson et al., 1988; Gladstone & Best, 1985; Kershner, 1985; Obrzut et al., 1981). Yet the *direction* of the dyslexics' underlying movement preference is obviously incompatible with Preilowski's (1975) neural model of mirror movements as it stands. If, however, we modify the model by hypothesizing anomalous organization of the ipsilateral motor pathways in our disabled readers, in addition to interhemispheric deficiency[1], it may extend to our findings. That is, parallel movements could occur if the ipsilateral pathway in dyslexics represents a *spatial match* of the movement direction projected in the contralateral pathway rather than the mirror-image representation proposed by Preilowski (1975).

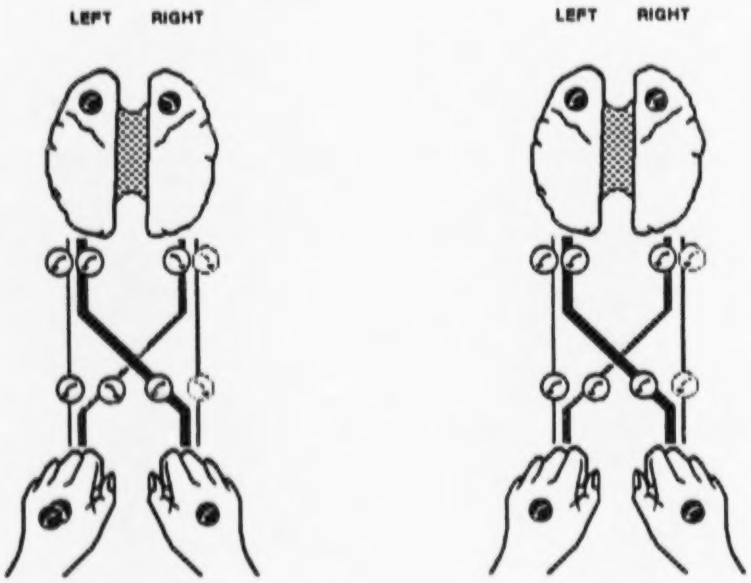


NORMAL SUBJECTS
Bidirectional Condition

NORMAL SUBJECTS
Unidirectional Condition

Figure 5a. Diagram of proposed influences from each hemisphere during bimanual performance by normal subjects (based on Preilowski, 1975). See text for detailed explanation.

Figure 5b illustrates our modified neural model of the unidirectional (parallel) bimanual movement preferences of our dyslexic subjects. As with the normal subjects, the left hemisphere is assumed to dominate in control of both hands, as illustrated by the darker pathways emanating from the left hemisphere than from the right hemisphere. Likewise, the contralateral pathways are assumed to be more influential than the ipsilateral pathways, as shown by their greater thickness. However, it is proposed that dyslexic children have deficient corpus callosum function, as illustrated by the lightly stippled corpus callosum and by the absence of callosally mediated influence of the left hemisphere over the right hemisphere. Furthermore, the ipsilateral pathways are presumed to convey a movement pattern that is spatially identical (rather than a mirror image, as in normals) to that conveyed by the contralateral pathway from the same hemisphere.



DYSLEXIC SUBJECTS
Bidirectional Condition

DYSLEXIC SUBJECTS
Unidirectional Condition

Figure 5b. Diagram of deficits in interhemispheric coordination and anomalous ipsilateral influences during bimanual performance by dyslexics. See text for detailed explanation.

Figure 5b thus indicates that unlike normal subjects, dyslexics should have greater difficulty with bidirectional (mirror) bimanual movements than with unidirectional (parallel) movements. In the former situation, conflicting commands would arrive at each hand from the contralateral versus the ipsilateral pathways. This would lead to left-hand performance deficits in particular, resulting from conflict between the contralateral pathway from the nondominant right hemisphere to the left hand versus the ipsilateral pathway from the dominant left hemisphere to the left hand. Unidirectional bimanual movements should be easier because each hand would receive spatially identical movement commands from its contralateral and ipsilateral pathways.

Furthermore, the combination of poor interhemispheric communication, anomalous ipsilateral representation and left hemisphere dominance for bimanual control (e.g., Kreuter et al., 1972; Wyke, 1971a, 1971b) could account for specific deficits in *left* hand performance especially when *bidirectional* (mirror) movements are required of dyslexics. This deviates from the pattern that Preilowski's model would predict for normal subjects—poor *left*-hand performance under requirements for *unidirectional* (parallel) bimanual movements (Figure 5a).

The possibility of a parallel-movement preference in some subpopulations, as opposed to the more typical mirror-movement preference, is supported by a study of bimanual drawing patterns in normal right- and left-handed children (Reed & Smith, 1961). Although most children chose to draw in rungs for two side-by-side "ladders" with simultaneous extensor (mirror) movements of the hands, a subset of the right-handers instead produced simultaneous left-to-right movements (parallel). Thus, the mirror preference does not appear to be universal, even in normal right-handed children. The proposal of anomalous ipsilateral representation of movement as an explanation of either our findings or those of Reed and Smith is, of course, post-hoc and quite speculative. However, as discussed earlier, no other existing model can account for a parallel movement preference. The present hypothesis would need to be tested in clinical populations known to have abnormal organization of both callosal and pyramidal pathways, such as callosal agenetics.

A number of additional questions remain to be addressed in future research. For example, the assumption that the bimanual pattern seen among the impaired readers is specific to that population, and not simply a phenomenon characteristic of general cerebral impairment, should be tested.[2] Also unanswered is whether the observed bimanual (bilateral) deficits reflect a general characteristic of interlimb coordination among dyslexics, which might be found even for movements of ipsilateral limbs.[3] Systematic developmental studies of bimanual coordination would also be useful. Although the analysis of covariance attempted to determine the effect of age on bimanual performance, the issue would be more adequately addressed through the inclusion of a younger control group. Our bimanual task presumably tapped interhemispheric coordination of frontal motor functions. It would be of interest to assess for anomalous interhemispheric coordination of cortical functions in other regions. Finally, while this task demonstrates deficits among dyslexics on a task that presumably requires interhemispheric interaction, the precise mechanism of the interhemispheric transfer anomaly remains to be determined. Recent research (Davidson et al., 1988) suggests that one such mechanism may be a deficit in the timing of interhemispheric transfer. The examination of the relation between interhemispheric transfer time and performance on the bimanual coordination task used in this study would provide valuable information on one possible mechanism for the interhemispheric transfer deficit.

There are also methodological limitations in our design that may warrant further study. Following Preilowksi (1972, 1975), we had predicted that dyslexics would have difficulty under bimanual conditions requiring different *ratios* of hand movements. Because we had not expected that the *direction* of hand movement would show group differences, we did not assess bimanual movement in all directions. The condition omitted would have required counterclockwise movement of both hands. Inclusion of this condition might further clarify the observed bias for parallel clockwise action of the hands. It should be noted, however, that parallel counterclockwise movements do not seem to pose a particular problem for dyslexics, given that several of them spontaneously produced such movements during reversals. A further design limitation was present in the NVF condition. While the apparatus occluded the subjects' view of the plotter screen, it did allow them to monitor their own hand movements visually. The elimination of all visual feedback in the NVF condition would provide a more adequate control condition.

In summary, our findings support the hypothesis that dyslexics manifest intermanual coordination deficits. The specific pattern of deficits that we found, however, was neither predicted nor previously reported. As a result, it is not readily explained by current models of interhemispheric collaboration. As an attempt to explain the observed pattern, we hypothesized that dyslexics may show both deficient interhemispheric processes and anomalous organization of the ipsilateral motor pathways. The novelty of the findings, however, indicates the need for further research.

ACKNOWLEDGMENT

This paper is based upon research conducted for the partial fulfillment of the Ph.D. degree requirements for the first author. The work and preparation of this paper were supported by March of Dimes Birth Defects Foundation grant No.12-165 awarded to the third author and by NICDH grant HD-01994 to Haskins Laboratories. We thank the following people for their invaluable help on this project: Barry H. Cohen for statistical and methodological support; Clifford Saron for technical support; Susan Leslie and Orit Batey for neuropsychological testing; Martha Denckla for her thoughtful comments on an earlier draft of the paper; Mary Beth Hickey and Nina Huza for their assistance in typing the paper; and particularly our subjects and their parents for their participation in the research.

REFERENCES

Badian, N. A., & Wolff, P. H. (1977). Manual asymmetries of motor sequencing in reading disability. *Cortex, 13*, 343-349.

Benton, A. L., & Spreen, O. (1969). *Benton-Spreen right-left orientation test*. Victoria, BC: University of Victoria.

Bender, L. (1938). A visual motor gestalt test and its clinical use. *American Orthopsychiatric Association Research Monographs, 3.*

Brinkman, J., & Kuypers, H. G. J. M. (1972). Split brain monkeys: Cerebral control of ipsilateral and contralateral arm, hand and finger movements. *Science, 176*, 536-539.

Davidson, R. J., Leslie, S. C., & Saron, C. (1988). *Reaction time measures of interhemispheric transfer time in reading disabled and normal children*. Submitted for publication.

Denckla, M. B. (1985). Motor coordination in dyslexic children: Theoretical and clinical implications. In F. Duffy & N. Geshwind (Eds.), *Dyslexia: A neuroscientific approach to clinical evaluation*. Boston: Little, Brown & Co.

Denckla, M. B. (1986). *Norman Geshwind Memorial Lecture*. Paper presented at the Annual Orton Society Meeting, Philadelphia.

Denckla, M. B., & Rudel, R. (1976). Rapid "automatized" naming (R. A. N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14*, 471-479.

Dennis, M. (1976). Impaired sensory and motor differentiation with corpus callosum agenesis: A lack of callosal inhibition during ontogeny? *Neuropsychologia, 14*, 455-469.

Drake, W. E. (1968). Clinical and pathological findings in a child with a developmental learning disability. *Journal of Learning Disabilities, 1*, 486-502.

Geschwind, N. (1975). The apraxias: Neural mechanisms of disorders of learned movement. *American Scientist, 63*, 188-195.

Gladstone, M., & Best, C. T. (1985). Developmental dyslexia: The potential role of interhemispheric collaboration in reading acquisition. In C. T. Best (Ed.), *Hemispheric function and collaboration in the child* (pp. 87-118). San Francisco: Academic Press.

Gross-Glenn, K., & Rothberg, S. (1984). Interhemispheric tactile transfer in dyslexics. *International Journal of Neuroscience, 24*, 23-35.

Haerer, A. F., & Currier, R. D. (1966). Mirror movements. *Neurology, 16*, 757-765.

Harris, A. J. (1958). *The Harris tests of lateral dominance*. New York: The Psychological Corp.

Hiscock, M., & Kinsbourne, M. (1980). Asymmetries of selective listening and attention switching in children. *Developmental Psychology, 16*, 70-82.

Kershner, J. R. (1985a). Ontogeny of hemispheric specialization and relationship of developmental patterns to complex reasoning skills and academic achievement. In C. T. Best (Ed.), *Hemispheric function and collaboration in the child* (pp. 327-360). San Francisco: Academic Press.

Kershner, J. R. (1985b, May). *Dual processing models of learning disability*. Paper presented at the conference on the Developmental Implications of Brain Lateralization, St. Catherines, Ontario, Canada.

Kershner, J. R., Henninger, P., & Cooke, W. (1984). Written recall induces a right hemisphere linguistic advantage for digits in dyslexic children. *Brain and Language, 21*, 105-122.

Kimura, D. (1977). Acquisition of a motor skill after left hemisphere damage. *Brain, 100*, 527-542.

Klicpera, C., Wolff, P. H., & Drake, C. (1981). Bimanual coordination in adolescent boys with reading retardation. *Developmental Medicine and Child Neurology, 23*, 617-625.

Kreuter, C., Kinsbourne, M., & Trevarthen, C. (1972). Are the disconnected hemispheres independent channels? A preliminary study of the effects of unilateral loading on bilateral finger tapping. *Neuropsychologia, 10*, 453-461.

Leslie, S. C., Davidson, R. J., & Batey, O. B. (1985). Purdue pegboard performance of disabled and normal readers: Unimanual versus bimanual differences. *Brain and Language, 24*, 359-369.

Levy, J. (1985). Interhemispheric collaboration: Single mindedness in the asymmetric brain. In C. T. Best (Ed.), *Hemispheric function and collaboration in the child* (pp. 11-31). San Francisco: Academic Press.

McGuinness, D. (1985). *When children can't learn*. New York: Basic Books, Inc.

Myklebust, H. R. (1968). Learning disabilities: Definition and overview. In H. R. Myklebust (Eds.), *Progress in learning disabilities* (Vol. I). New York: Grune and Stratton.

Nass, R. (1985). Mirror movement asymmetries in congenital hemiparesis: The inhibition hypotheses revisited. *Neurology, 35*, 1059-1062.

Obrzut, J., Obrzut, A., Hynd, G., & Pirozzolo, F. (1981). Effect of directed attention on cerebral asymmetries on normal and learning disabled children. *Developmental Psychology, 17*, 118-125.

Orton, S. (1937). *Reading, writing and speech problems in children*. New York: Norton.

Owen, F., Adams, P. A., Forrest, T., Stolz, L. M., & Fisher, S. (1971). Learning disorders in children sibling studies. *Monographs of the Society for Research in Child Development, 36*. Chicago: University of Chicago Press.

Preilowski, B. (1972). Possible contributions of the anterior forebrain commissures to bilateral motor coordination. *Neuropsychologia, 10*, 267-277.

Preilowski, B. (1975). Bilateral motor interaction: Perceptual motor performance of partial and complete "split brain" patients. In K. J. Zulch, O. Creutzfeldt, & G. C. Galbraith (Eds.), *Cerebral localization* (pp. 115-132). Berlin: Springer.

Reed, G. F., & Smith, A. C. (1961). Laterality and directional preference in a simple perceptual motor task. *Quarterly Journal of Experimental Psychology, 13*, 122-124.

Rudel, R. G. (1985). Hemispheric asymmetry and learning disabilities: Left, right or in between? In C. T. Best (Ed.), *Hemispheric function and collaboration in the child* (pp. 275-308). San Francisco: Academic Press.

Sklar, B., Hanley, J., & Simmons, W. W. (1973). A computer analysis of EEG spectral signatures from normal and dyslexic children. *IEEE Transactions on Biomedical Engineering, 20*, 20-26.

Sperry, R. W., Gazzaniga, M. S., & Bogen, J. E. (1969). Interhemispheric relationships: The neocortical commissures; syndromes of hemisphere disconnection. In P. J. Vinken & G. W. Bruyn (Eds.), *Handbook of clinical neurology* (Vol 4, pp. 273-290). Amsterdam: North-Holland.

Spreen, O., & Benton, A. L. (1977). *Neurosensory center comprehensive examination for aphasia*. Victoria, British Columbia: Neuropsychology Laboratory, Department of Psychology, University of Victoria.

Tiffin, J. (1968). *Purdue Pegboard examiners' manual*. Chicago: Science Research Associates.

Tomaino, C. A. (in press). Dyslexia and abnormal motor movements: A review of the literature. *Journal of Educational Neuropsychology*.

Touwen, B. C. L., & Prechtl, H. F. R. (1970). The neurological examination of the child with minor nervous dysfunction. *Clinics in Developmental Medicine, 38*. London: Heinemann.

Vaughn, H. G., & Costa, L. D. (1962). Performance of patients with lateralized cerebral lesions. II: Sensory and motor tests. *Journal of Nervous and Mental Disease, 134*, 237-243.

Wechsler, D. (1974). *Manual for the Wechsler intelligence test for children-revised*. New York: Psychological Corporation.

Wolff, P. H., Cohen, C., & Drake, C. (1984). Impaired motor timing control in specific reading retardation. *Neuropsychologia, 22*, 587-600.

Woodcock, R. E. (1973). *Woodcock reading mastery tests*. Circle Pines, MN: American Guidance Service Inc.

Woodcock, R. W. (1976). *Goldman-Fristol-Woodcock auditory skills battery*. Circle Pines, MN: American Guidance Services, Inc.

Wyke, M. (1971a). The effect of brain lesions on the performance of bilateral arm movements. *Neuropsychologia, 9*, 33-42.

Wyke, M. (1971b). The effect of brain lesions on the learning performance of a bi-manual coordination task. *Cortex, 7*, 59-72.

Yeni-Komshian, G. H., Isenberg, D., & Goldberg, H. (1974). Cerebral dominance and reading disability: Left visual field deficit in poor readers. *Neuropsychologia, 12*, 131-139.

FOOTNOTES

[*] *Developmental Psychology*, 1989, *25*, 236-246.

[†] Department of Psychology, New York Hospital-Cornell Medical Center, now at Department of Clinical Psychology, Institute of Living, Hartford, CT.

[††] Department of Psychology, Wesleyan University.

[†††] Department of Psychology, University of Wisconsin-Madison.

[1] Our proposal assumes both inhibitory and facilitative interhemispheric influences during the performance of bimanual coordination tasks.

[2] We thank Steven Mattis for suggesting this possibility to us.

[3] Our thanks to Paula Tallal for making this point to us.

Expressive Microstructure in Music: A Preliminary Perceptual Assessment of Four Composers' "Pulses"*

Bruno H. Repp

According to a provocative theory set forth by Manfred Clynes, there are composer-specific cyclic patterns of (unnotated) musical microstructure that, when discovered and realized by a performer, help to give the music its characteristic expressive quality. Clynes, relying mainly on his own judgment as an experienced musician, has derived such personal "pulses" for several famous composers by imposing time and amplitude perturbations on computer-controlled performances of classical music and modifying them until they converged on some optimal expression. To conduct a preliminary test of the general music lover's appreciation of such "pulsed" performances, two sets of piano pieces by Beethoven, Haydn, Mozart, and Schubert, one in quadruple and the other in triple meter, were selected for this study. Each piece was synthesized with each composer's pulse and also without any pulse. These different versions were presented in random order to listeners of varying musical sophistication for preference judgments relative to the unpulsed version. There were reliable changes in listeners' pulse preferences across different composers' pieces, which affirms one essential prerequisite of Clynes' theory. Moreover, in several instances the "correct" pulse was preferred most, which suggests not only that these pulse patterns indeed capture composer-specific qualities, but also that listeners without extensive musical experience can appreciate them. In other cases, however, listeners' preferences were not as expected, and possible causes for these deviations are discussed.

INTRODUCTION

It is widely agreed that, in the performance of notated music, particularly Western art music from the 18th and 19th centuries, literal reproduction of the written score does not result in a very satisfying experience. The notation, in its rigid subdivision of note values and its lack of dynamic instructions for individual notes, omits many of the composer's intentions. To make the performance interesting, expressive, and musically satisfying, much variation not conveyed by the written score must be introduced by the performer. This variation may be relatively small (but nevertheless important) with respect to timing of note onsets, relative note durations, and pitch, where the score makes specific prescriptions; it may be much larger with respect to relative note intensities and other aspects of articulation, where the notation provides few constraints. All this variation constitutes the "expressive microstructure" (Clynes, 1983) of a performance. The principles musicians follow in generating this microstructure are not well understood and are only to a very limited degree made explicit in traditional music instruction. A great interpreter's skills may seem mysterious and beyond explanation. The psychology of musical expression is still in its infancy, but there is a rich lode waiting to be mined.

Performers' expressive devices may be divided roughly into four categories according to their origin and motivation:

(1) First, there are those variations that are contingent on local musical events, such as the structure of phrases, the shape of the melodic line, and the harmonic progression. These deviations emphasize structural properties and thus help realize some of the expressive potential inherent in a musical composition. They are relatively amenable to introspection, instruction, and scientific investigation, as they are likely to follow certain *general* rules that are observed, more or less, by all competent musicians. They also apply to a wide range of music by different composers. Progress towards understanding these rules has been made by several researchers (e.g., Bengtsson & Gabrielsson, 1980; Clynes, 1983; Gabrielsson, Bengtsson, & Gabrielsson, 1983; Shaffer, 1981; Shaffer, Clarke, & Todd, 1985; Sundberg, Frydén, & Askenfelt, 1983; Sundberg & Frydén, 1985; Todd, 1985).

(2) Second, there are those aspects of expressive microstructure that reflect a performer's understanding of a composer's individual characteristics as they are conveyed in his musical oeuvre as a whole. These aspects are much more elusive and difficult to investigate, as they seem to be the province of truly gifted interpreters whose performances have the "ring of authenticity." Nevertheless, one very intriguing attempt to understand these *composer-specific* principles of expressive microstructure has been made (Clynes, 1969, 1983, 1986, 1987).

(3) Third, there are those deviations that reflect an *individual performer's* style and, perhaps, mannerisms. It is these characteristics that shape distinctive interpretations, and that enable the experienced listener to recognize certain artists from their performances. To a large extent, they may reflect the specific use a performer makes of the rules mentioned under (1) and (2), but there may also be genuinely personal patterns of expression. This is uncharted territory for the music psychologist.

(4) Finally, there are various *piece-specific* factors which may derive from notated dynamic instructions, performance conventions, explicit "programs," and also from motoric limitations in executing difficult passages.

In general, therefore, the expressive microstructure of a musical performance reflects general, composer-specific, performer-specific, and piece-specific factors. The present investigation is concerned with only one of these factors, the composer-specific one, based on the theory of expressive microstructure developed by Clynes (1983, 1986).

The idea that different composers have different personalities that are conveyed in their music and that need to be understood and expressed by performers is very old, of course; it is part of the general lore of music history, performance, and criticism. Until recently, however, these composer-specific characteristics, especially as they go beyond the printed score, have eluded quantification. Extending the method of "accompanying movements" (*Begleitbewegungen*) developed by the German philologist Eduard Sievers

(1924), the German musicologist Gustav Becking (1928) made an interesting, though still very rough and introspective, attempt at more precise description by observing his own movements as he conducted along while listening (more precisely, as he "let himself be conducted by the music"). He found characteristic movement trajectories of his conducting index finger for music by different composers, which he interpreted as reflections of the composers' personality and general attitude to the world. However, he did not relate these dynamic patterns to performance microstructure, which at the time was only beginning to be measured objectively (e.g., Hartmann, 1932; Seashore, 1938/1967). It remained for Manfred Clynes, a noted inventor, neuroscientist, and musician, to achieve a more precise quantification of these dynamic characteristics, thus making them amenable to scientific investigation.

Clynes' initial step, some two decades ago, consisted in going from movement to pressure, using a pressure-sensitive recording device called the sentograph (Clynes, 1969). At that time, Clynes asked several prominent musicians (including Pablo Casals and Rudolf Serkin) to "conduct" by rhythmically pressing the sentograph with their finger about once a second while mentally rehearsing specific compositions of several different composers. The resulting periodic pressure curves were averaged to yield a single "pulse" shape for each musician and each composition. It emerged that these pulse shapes were remarkably similar across several different compositions (both slow and fast) by the same composer, and also across different musicians imagining the same piece. They were very different for different composers, however. These rather limited but striking observations confirmed Becking's idea of composer-specific "pulses" that can be externalized as movement patterns.

The second important step taken by Clynes more recently were his investigations into how these pulses might be conveyed in the actual microstructure of music performance (Clynes, 1983, 1986, 1987). The central assumption underlying this effort is that composer-specific pulses are not restricted to people's musical thought and accompanying movements, but that they can be *physically instantiated* in the musical sound pattern, with benefits for the listener. These pulses are defined as patterns of systematic deviations from the notated relative durations[1] and (usually unspecified, hence nominally equal) relative loudnesses of the notes within a time unit (e.g., one bar). Thus there are two independent components in each pulse pattern. According to Clynes' theory, the pattern applies throughout a composition; that is, it repeats itself in a cyclical fashion (e.g., bar by bar) from beginning to end. The motivation for this requirement lies in earlier ideas of Clynes (see Clynes & Walker, 1982) concerning "time form printing": Recurring time-amplitude patterns are assumed to set up expectancies in the central nervous system for the patterns to continue. The pulse, despite its rigid recurrence, is assumed to imbue music with "living qualities" that specifically reflect the composer's personality and that may also enhance the expressiveness of his characteristic melodic contours. Clynes (1983) has likened the pulse to such other individual motor characteristics as gait, handwriting, and speech. As a personal "style of movement," it is assumed to apply to all works of a composer.

Rather than measuring the time-amplitude patterns of actual performances by great artists, Clynes has developed composer-specific pulses by means of computer synthesis, relying mainly on his own judgment as an exceptionally sensitive musician. He has developed software[2] that enables him to enter a musical score into the computer and to specify a pattern of relative durations and intensities, which then determines the exact values of the notes within each time unit (chosen to coincide with a notated time unit close to one second in duration). Depending on the time signature of the composition, the time units are divided into either three or four (sometimes two) subunits, with separate corresponding pulse patterns. Notes lasting longer than one subunit are assigned the sum of the component durations and the amplitude of the first component they occupy.[3] The pulse is often also implemented at a second, higher level, with three or four time units as subunits. Thus, for example, in a fast piece in 3/4 measure, the basic 3-pulse would comprise one bar and the higher-level pulse (usually a 4-pulse, reflecting the phrase structure) would comprise four bars. In a slower piece in 3/4 measure, on the other hand, there would be a basic 4-pulse for the sixteenth-notes within each quarter-note and a higher-level 3-pulse comprising one bar.

By experimenting with many different compositions and pulses and by carefully listening to the results (see Clynes, 1983), Clynes has arrived at what he considers appropriate pulse specifications for a number of famous composers. Four of these—the Beethoven, Haydn, Mozart, and Schubert pulses—are illustrated in Table 1, each in a quadruple-meter, a triple-meter, and (if used in the present study) a duple-meter version. In each pulse specification, the first line indicates the relative durations (in percent) of successive notes (i.e., onset-to-onset intervals) of equal nominal durations, and the second line indicates their relative amplitudes (in linear proportions). The duration and amplitude components do not follow the same pattern; they are independent parameters. In the Beethoven 4-pulse, for example, the rank order of the four beats is 4-1-3-2 in terms of duration, but 1-3-4-2 in terms of amplitude. The 4-pulse, 3-pulse, and 2-pulse patterns for the same composer are related (see Clynes, 1987). For a detailed discussion of the characteristic features of these composers' 4-pulses and their interpretation in terms of dynamic qualities, see Clynes (1983, pp. 134-135).

TABLE 1. Pulse specifications in quadruple, triple, and duple meter for four composers. The first line indicates relative note durations (i.e., onset-to-onset intervals) in percent; the second line represents relative amplitudes on a linear scale. These pulses are the ones actually applied in the present materials (Clynes, p.c.); the triple- meter pulses differ from the preliminary specifications of Clynes (1983, 1986). Note that triple-meter timing pulses are not normalized, so that slight tempo changes result, and that all amplitude pulses result in various degrees of attenuation.

Composer	Quadruple meter				Triple meter			Duple meter	
Beethoven	106	89	96	111	105	88	107	(not used)	
	1.00	0.39	0.83	0.81	1.00	0.46	0.75		
Haydn	108	94	97	102	108	95	103	100	97
	1.00	0.42	0.68	1.11	1.03	0.35	0.60	1.00	0.65
Mozart	105	95	105	95	106.5	102.5	97.5	100	100
	1.00	0.21	0.51	0.23	0.78	0.25	0.30	1.00	0.51
Schubert	98	115	99	91	103	114	97.5	(not used)	
	1.00	0.65	0.40	0.75	0.92	0.45	0.72		

Clynes' pulse patterns reflect the musical insight and extensive efforts of one individual; they quantify a subjective experience. It remains to be shown that the pulse patterns chosen generalize to other listeners; perhaps they require exceptional musical sensitivity to be appreciated at all, and perhaps they are entirely idiosyncratic. At public occasions and on recordings accompanying some of his publications (Clynes, 1983, 1985, 1987), Clynes has presented many examples of music synthesized with appropriate composers' pulses, though their impressions on listeners were never documented in a formal way. Some recent demonstrations (Clynes, 1987) have included the same piece synthesized with several different composers' pulses, including the "correct" one. In the fall of 1985, the author received such a demonstration tape from Clynes containing the final movement of a Haydn Piano Sonata synthesized with six different composers' pulses. A small informal group of listeners agreed that the "Haydn pulse" version indeed was the most pleasing of the lot, whereas several other versions were perceived as rhythmically irregular and/or inappropriately accented. That one set of time-amplitude irregularities sounded "normal" while others sounded uneven was an interesting experience; it suggested that even the average listener could appreciate the appropriate expressive microstructure. However, it could have been that the Haydn pulse was simply perceptually more regular than the others on psychoacoustic or general musical grounds. If so, it might also have been preferred if the piece had been by Mozart or Beethoven. For a rigorous test of the prediction that, for each of several composers' music, that composer's pulse should be perceived as more appropriate than any other composer's pulse, a set of balanced materials was required in which each composer's piece was performed with each composer's pulse. The author, not having the necessary synthesis capabilities at the time, approached Clynes, who agreed to generate two such sets of materials for the experiments described below.

The experiments tested the following *five predictions*:

(1) Listeners' preferences for individual pulses should vary as a function of composer. That is, with composer and pulse as orthogonal factors in the experimental design, a statistical interaction should be obtained. Conversely, the null hypothesis was that, while some pulses may be preferred over others, these preferences would hold regardless of composer. Rejection of the null hypothesis would support the general claim of Clynes' theory that different composers require different pulses.

(2) For each composer listeners should prefer the "correct" pulse over all others. This hypothesis concerns the validity and generality of the specific pulse patterns devised by Clynes.

(3) Listeners should also prefer the correct pulse over a literal rendition of the score that has no pulse applied to it. Such "neutral" versions were included in the present materials. Clynes' theory admits the possibility that other pulses besides the correct one are preferred over a neutral version, but it seems essential that the correct pulse be perceived as an improvement over no pulse at all.

(4) The correct pulse should be preferred in all compositions by the same composer. This hypothesis, which concerns the generality of the pulses across each composer's oeuvre, could be tested only to a very limited extent in the present experiments, due to the small sample of materials.[4]

(5) The degree to which listeners are able to appreciate the correct pulse and, more generally, expressive microstructure in music, may be a function of their musical training and experience with classical music. That is, the strongest support for Clynes' theory might come from the most sophisticated listeners, though the responses of less experienced subjects were also of interest in this study.

EXPERIMENT 1

Experiment 1 used a set of pieces with even (2/4) time signature, to which Clynes' 4-pulses (and 2-pulses) had been applied. The experiment was conducted in two versions, referred to as 1a and 1b. It was initially assumed that the performances with the "correct" pulse should be the most satisfying on general musical grounds; thus, the subjects in Experiment 1a were instructed to indicate simply how much they liked each version. It was later pointed out to the author by Clynes (personal communication) that the correct pulse may not always be the most pleasing one, especially to the less experienced listener; for example, the rough Beethoven pulse may be less pretty than the gentle Haydn pulse when applied to a Beethoven piece, but it nevertheless characterizes the composer better. Some data collected independently by Thompson (submitted) with some of the same materials (the Beethoven piece) reinforce this point. Therefore, the experiment was repeated with modified instructions that emphasized composer-appropriate expression as the criterion for subjects' judgments (Experiment 1b). In addition, the quality of sound reproduction was improved in that experiment.

Methods

Subjects

Experiment 1a. Sixteen unpaid volunteers (7 women and 9 men, including the author) served as listeners.[5] All were investigators or graduate students in psychology or related areas at Haskins Laboratories or Yale University, mostly 25-35 years old, though two were in their forties. Most of them had received a musical education, and the majority were active amateur musicians and/or spent much time listening to classical music. A few musically less experienced subjects were included to extend the range for purposes of correlational analyses. According to a questionnaire filled out by the subjects at the end of the experiment, they had had between 0 and 25 years (summed over all instruments) of formal instruction on various instruments (including piano, violin, cello, guitar, saxophone, and voice), spent between 0 and 23 hours a week playing their instruments, and listened to serious music 0 to 15 hours per week.

Experiment 1b. Thirteen different subjects participated here; their age and gender were not recorded. Four were faculty members at Trinity College (Hartford, CT); most of the others were undergraduate students enrolled in a music course there; one was a researcher at Haskins Laboratories. They included a professional pianist, a composer, an accomplished organist, several musical amateurs, and a few musically naive individuals. According to the questionnaire, they had had between 0 and 23 years of formal musical instruction, played their instruments from 0 to 14 hours per week, and listened to classical music between 0 and 45 (!) hours per week.

Materials

Four piano compositions, one each by Beethoven, Haydn, Mozart, and Schubert, were selected by the author. All were in 2/4 time signature and had a fast tempo, except for the Schubert, which was both longer and slower than the other pieces. They were:

Beethoven: last movement (Presto) of the Piano Sonata in F major, op.10, No. 2 (complete, without repeats);

Haydn: last movement (Presto) of the Piano Sonata in F major, HV XVI/23 (complete, without repeats);

Mozart: exposition (i.e., the first part) of the last movement (Allegro) of the Piano Sonata in C major, K. 279 (without repeat);

Schubert: Moment Musical No. 4 (Moderato) in c-sharp minor, op. 94, D. 780 (complete, without repeats; the contrasting middle section is in D-flat major).

The initial bars of these pieces are shown in Figure 1.

All computer performances were generated by Manfred Clynes at the Music Research Institute of the New South Wales State Conservatorium of Music in Sydney, Australia, using his special software developed there. These computer realizations, in contrast to earlier recorded examples (e.g., Clynes, 1983), included all the notes, not just the melody. Each piece was recorded on cassette tape (with Dolby B noise reduction) in five versions: without any pulse (the "neutral" version) and with each of the four composers' pulses. The pulse patterns used are shown in Table 1. They were applied at two hierarchical levels, the lower (main) level comprising four sixteenth-notes and the higher level comprising two or four quarter-notes (i.e., one or two bars). In a number of instances, an "attenuation factor" was applied to the timing pattern at the higher level, which changed the depth of modulation while keeping the specified relationships constant.[6] The pulses applied, together with these attenuation factors, are shown in Table 2. The choice of attenuation factors reflects Clynes' artistic judgment. In addition, a few manual adjustments (such as appropriate ritardandos and micropauses) were made by Clynes to enhance the general musical quality of the performances; these were common to all versions, including the neutral one.

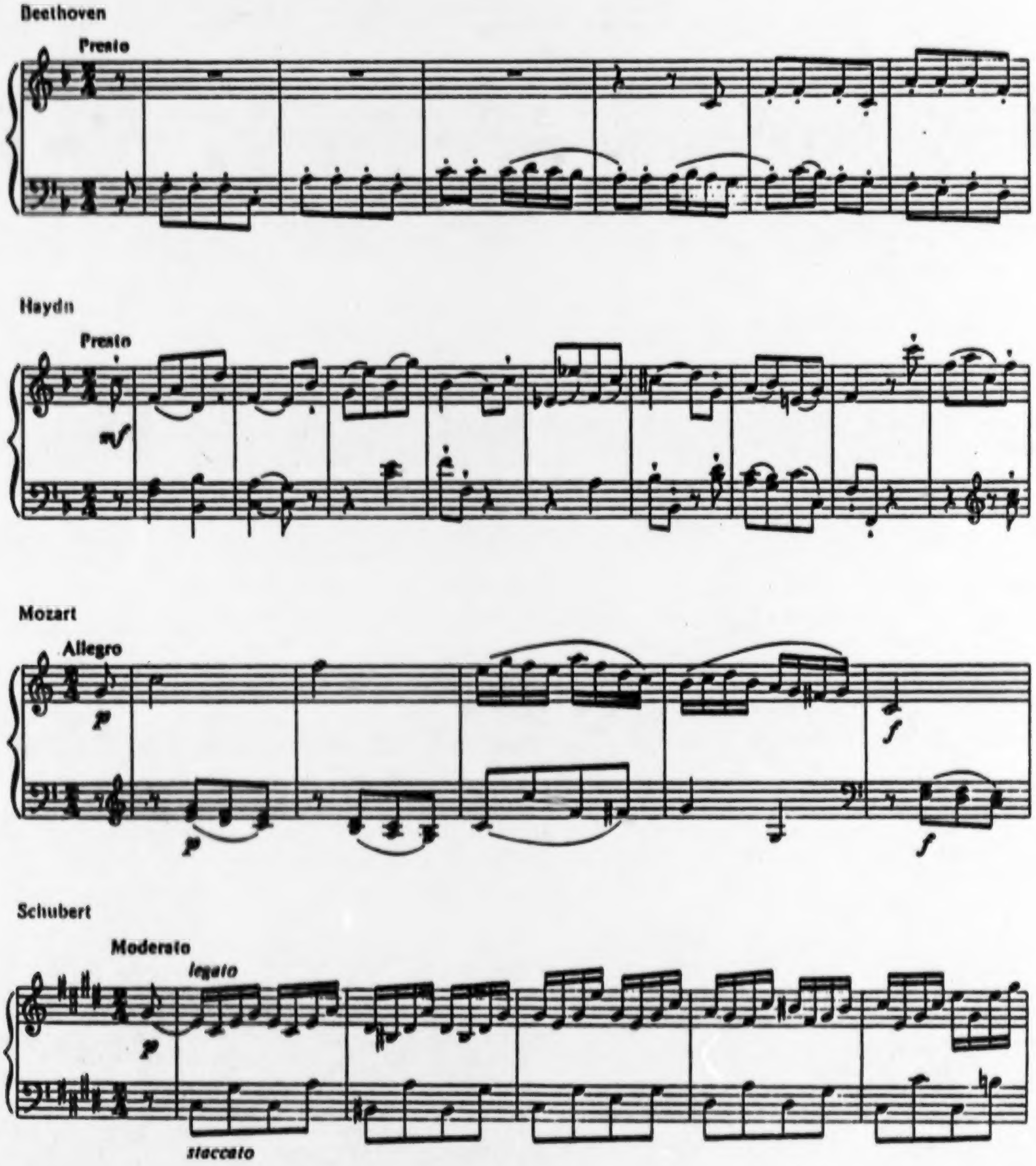


Figure 1. Initial bars of the four pieces used in Experiment 1.

TABLE 2. Hierarchical pulse patterns (see Table 1) and attenuation factors for the timing pulses used in synthesizing the pieces of Experiment 1 (Clynes, p.c.).

Composer	Lower level	Higher level
Beethoven	4-pulse (1.1)	4-pulse (0.6)
Haydn	2-pulse (1.0)	4-pulse (1.0)
Mozart	2-pulse (1.0)	4-pulse (1.0)
Schubert	4-pulse (1.0)	4-pulse (0.15)

The Mozart and Schubert pieces were reproduced on a Roland MKS-20 piano sound module, the Beethoven on a Prophet 2000 synthesizer with an "early 19th century pianoforte" sound.[7] Since the Haydn piece was already available on the demonstration tape mentioned in the introduction, it was not resynthesized by Clynes. Since computer-generated sinusoids were used in its synthesis, it had a more artificial sound quality; on the other hand, since it was free from the envelope constraints of a simulated piano, it included "predictive amplitude shaping" (see Clynes, 1983) of individual tones in all versions. Unfortunately, the demonstration tape did not include a "neutral" version of the Haydn, but versions with two other composers' pulses instead; so the version with the Schumann pulse was substituted for the neutral version.

An experimental tape was generated by dubbing from the master tape. The order of the four pieces was: Haydn, Mozart, Schubert, Beethoven. The five versions of each piece occurred in succession, separated by approximately 10 seconds of silence. The first version was always the neutral version, and the four pulsed versions followed, in an order that was different for each piece and represented a 4 x 4 Latin square design.

Procedure

Experiment 1a. Some subjects were tested at Haskins Laboratories, others at Yale University. The subjects were seated in a quiet room and listened to the music over a single loudspeaker connected to a high-quality cassette deck (at Haskins) or over the stereo loudspeakers of a portable cassette recorder (at Yale). Each subject received detailed written instructions, including the following:

> You will hear a number of short piano pieces reproduced by a synthesizer. Each of these pieces will be played in *five versions*. The first, "neutral" version will be an almost literal rendition of the score, with only a few essential deviations from the written notation. The four subsequent versions will include additional adjustments in the timing and relative loudness of the individual notes. Each version will follow a different pattern, and some versions may be more successful than others. That is, while some versions may sound *more lively, expressive, and idiomatic* than the neutral version, others may sound slightly *distorted* in rhythm and accent pattern. (One way of thinking about these versions is that they represent performances by four different pianists of equal technical competence but different musical sensibilities.) Your task will be to indicate whether or not each particular version constitutes an improvement over the neutral version.
>
> Your judgments are to be made on an 11-point *rating scale* ranging from -5 to 5. The neutral version of any piece is identified with the midpoint (zero) on this scale, so you don't have to give any rating after hearing this first version. For

> each of the following four versions, you should enter your rating of that version, circling *one* number on the rating scale. Please, make your judgment *relative to the neutral version*: Assign a positive rating if you *prefer* the version just heard over the neutral version (because it seems more lively, expressive, and idiomatic) and a negative rating if you prefer the neutral version over it (because it seems irregular, distorted, and unidiomatic).

The answer sheet illustrated the rating scale and listed the numbers -5 through +5 for each version, one of which the subjects were to circle. For the neutral version, the subjects were to indicate instead their familiarity with the piece by circling "very," "moderately," "barely," or "not at all." The composer of each piece was named on the answer sheet. The whole session lasted about one hour.

Experiment 1b. One possible concern with Experiment 1a (Clynes, personal communication) was that the sound reproduction equipment was not optimal; in particular, the portable recorder used at Yale did not have the Dolby B option, which distorted the amplitude profiles of the pulses. In Experiment 1b, the tapes were played with Dolby B noise reduction, either at Trinity College or on the subject's home stereo equipment. More importantly, there was a change in instructions, as follows:

> While some versions may sound *expressive in a way that seems to befit the composer*, others may sound less convincing or even ridiculous. (...) Your task will be to indicate whether or not each particular version constitutes an improvement over the neutral version in that it captures some of the composer's characteristic expression—that is, whether each version sounds more or less "Beethovenian" or "Haydnian" or "Mozartian" or "Schubertian" than the neutral version.

Results and Discussion

Experiment 1a

One initial question was whether the subjects would be able to make consistent judgments at all. Not only were they not professional musicians, but also musical tastes are often said to be highly variable and idiosyncratic. The statistical analysis dispelled these fears. In a repeated-measures analysis of variance (ANOVA) on the subjects' ratings with two crossed factors, composer and pulse, the main effect of pulse was highly significant, $F(3,45) = 10.09$, $p < .0001$, which indicates that there were consistent preferences for certain pulses over others. Moreover, and much more importantly, there was a significant composer by pulse interaction, $F(9,135) = 4.43$, $p < .0001$. Thus the pattern of pulse preferences varied reliably with composers (pieces), which supports the first of the five predictions made in the introduction. There was no significant main effect of composer.

The average ratings are represented in Table 3, with the italicized numbers in the diagonal representing the "correct" composer-pulse combinations. It is evident that in three of the four pieces (Haydn, Mozart, Schubert) the correct pulse was indeed the one preferred most, although in the Mozart piece the Mozart and Haydn pulses were tied. Moreover, in each of these three cases the correct pulse received positive ratings that clearly exceeded the neutral baseline (subjects tended to be conservative in their ratings, using mostly the range between -3 and +3), although this difference is not meaningful in the case of the Haydn piece, whose "neutral" version really had an inappropriate pulse in it. Only the Beethoven piece produced disappointing results. For it, the listeners clearly preferred the Haydn pulse (which perhaps can be explained by the somewhat Haydn-like quality of that movement—Clynes, personal communication), whereas the Beethoven pulse received a negative average rating. Thus, in three out of four cases the results provide support for predictions 2 and 3, which concern the adequacy of the composers' pulses.

TABLE 3. Average ratings of the computer performances in Experiment 1a.

Composer	Pulse			
	Beethoven	Haydn	Mozart	Schubert
Beethoven	*-1.13*	2.25	-1.13	-0.13
Haydn	0.06	*1.69*	0.69	-0.69
Mozart	0.38	1.56	*1.56*	-0.75
Schubert	-0.25	1.19	-0.38	*1.94*

A look down the columns of Table 3 reveals that the Beethoven pulse was not liked much in any piece, whereas the Haydn pulse was liked in *all* compositions. Listeners' judgments of the Beethoven and Haydn pulses thus did not seem to vary much across compositions. This was confirmed by conducting an ANOVA on these two pulses only: The pulse main effect was highly significant, *F*(1,15) = 28.73, *p* = .0001, indicating that the Haydn pulse was preferred over the Beethoven pulse, but the composer by pulse interaction fell short of significance. Thus it could be argued that listeners' preference for the Haydn pulse in the Haydn piece was due to a general preference, not to composer-specific factors.

The situation was different for the Mozart and Schubert pulses, however. Although the Mozart pulse was tied with the Haydn pulse as the preferred pulse for the Mozart piece, it was liked better in the Mozart than in any other piece. Even more strikingly, the Schubert pulse was liked *only* in the Schubert, being mildly disliked with all other composers. The reliability of this interaction was confirmed in an ANOVA on this half of the data: There were no significant main effects, but a significant composer by pulse interaction was obtained, *F*(3,45) = 6.90, *p* = .0006. This part of the data, therefore, provides unequivocal support for Clynes' Mozart and Schubert pulse patterns, as well as for the present subjects' ability to appreciate them.

The data were analyzed in yet another way, by first averaging the ratings of the three incorrect pulses for each piece and then entering the data into a 4 x 2 ANOVA with composer and correct/incorrect pulse as factors. There was a significant main effect of pulse, *F*(1,15) = 10.96, *p* = .0048, which confirms that, overall, correct pulses received higher ratings than incorrect pulses. However, since this was not true in the Beethoven piece, there was also a highly significant composer by pulse interaction, *F*(3,45) = 9.43, *p* = .0001. In a separate analysis of the correct pulse ratings only, the grand mean was significantly larger than zero, *F*(1,15) = 8.67, *p* = .0101, which confirms that, overall, correct pulses were preferred over no pulse at all, again with the exception of the Beethoven, which resulted in a significant effect of composer, *F*(3,45) = 8.64, *p* = .0001.

Prediction 4 (generality of pulses) could not be addressed within the present experiment, since only a single piece of each composer was used. (However, see footnote 4.) Therefore, we turn to prediction 5, concerning the role of subjects' musical experience. As pointed out earlier, the sixteen subjects represented a rather wide range of musical experience, with only the professional level missing. For each subject, a measure of the degree to which he or she appreciated the correct pulses (a "pulse appreciation index" or PAI) was computed by subtracting the average rating of the 12 incorrect composer-pulse combinations from the average rating of the 4 correct combinations. (Thirteen of the sixteen subjects had positive indices.) Three indices of musical experience were available: number of years of musical education, added up over all instruments studied; number of hours spent playing music per week; and number of hours spent listening passively to serious music. Correlations were computed between all three measures and the PAI. The first and third measures did not correlate at all with the PAI,[8] whereas the second (active playing hours) showed a small positive correlation ($r = 0.36$) that, however, fell short of significance. Thus there was only a hint that listeners with greater *active* musical experience might show greater appreciation for composers' pulses.

Two subsidiary questions were addressed in similar correlational analyses. First, it was noted that some subjects gave mostly positive ratings, while others gave mostly negative ratings. A subject's overall average rating is an indirect measure of the extent to which he or she liked the neutral version, relative to the pulsed versions. The correlations between subjects' average ratings and the three measures of musical experience were mildly negative and nonsignificant. If musically more experienced subjects liked the neutral version somewhat more than did musically inexperienced subjects, a reason for this may have been that the more sophisticated subjects expected too much from the pulsed versions as "performances." This point will be taken up in the general discussion. Second, a familiarity index was computed for each subject by averaging his or her familiarity ratings for the four pieces (very = 3, moderately = 2, barely = 1, not at all = 0). On the whole, the subjects were not very familiar with the pieces: Twelve subjects had average ratings of 1 or less, and only two subjects (the author being one) were at least moderately familiar with all of them. No piece was much more familiar than the others. The correlation between the familiarity index and the PAI was -0.06, indicating that familiarity did not increase the appreciation of the correct pulses.

Experiment 1b

Experiment 1b, it will be recalled, presented the same test to a comparable group of subjects with new instructions that emphasized composer-specific expression, and with Dolby B sound reproduction. It is clear from Table 4, however, that these combined changes did not lead to results that were more favorable to Clynes' theory—on the contrary. In the Beethoven piece, the Beethoven pulse was disliked even more than previously, in striking contrast to findings of Thompson (submitted), which showed a preference for the Beethoven pulse in the very same materials under similar instructions. (The reason for this discrepancy is not clear.) The Haydn pulse, which listeners had liked in Experiment 1a, was not preferred significantly over the neutral version, though it still came out on top. In the Haydn piece, the previous preference for the appropriate pulse was no longer evident, and the listeners showed a preference for the Beethoven pulse instead. In the Mozart piece, where in Experiment 1a the Haydn and Mozart pulses had been preferred equally, the Haydn pulse was now preferred over the Mozart pulse. Only in the Schubert piece did the Schubert pulse still come out on top, but it was virtually tied with the Haydn pulse and rated only slightly above the neutral version.

A look down the columns of Table 4 is only slightly more encouraging. At least it was still the case that the Mozart pulse was liked best in Mozart, and the Schubert pulse in Schubert. However, the Beethoven pulse was liked best in Haydn (and was strongly disliked in Beethoven), and the Haydn pulse was liked best in Mozart.

TABLE 4. Average ratings of the computer performances in Experiment 1b.

Composer	**Pulse**			
	Beethoven	**Haydn**	**Mozart**	**Schubert**
Beethoven	*-3.08*	0.15	-1.31	-1.69
Haydn	1.62	*0.15*	0.08	0.08
Mozart	0.08	1.54	*0.62*	-0.46
Schubert	-0.31	0.69	0.00	*0.77*

The data were also less consistent in this experiment. The ANOVA showed a marginally significant main effect of pulse, $F(3,36)$ 3.07, $p = 0.0399$, due to a general preference for the Haydn pulse, and a marginally significant composer by pulse interaction, $F(9,108)$ 1.97, $p = 0.0492$. In contrast to Experiment 1a, there was a highly significant main effect of composer, $F(3,36)$ 7.16, $p = 0.0007$, which reflected the strongly negative ratings for the Beethoven piece. Moreover, correlational analyses of average ratings and individual PAI values (5 positive, 8 negative) in relation to indices of musical experience and familiarity ratings revealed not a single significant correlation. Thus, the more experienced listeners did not give responses that were more in conformity with Clynes' theory. At the same time, the present group of subjects gave higher familiarity ratings than that of Experiment 1a, so the results cannot be attributed to general inexperience.

In summary, this experiment showed that instructions to rate the composer-specific expressive quality of the computer performances did not increase subjects' appreciation of composers' pulses. If anything, responses were more variable, which suggests that the instructions proved confusing. While it is relatively easy to express a simple preference for one or another performance, it is much more difficult (and perhaps presumptuous) to make judgments about the "Beethovenian" or "Mozartian" quality of expression. The subjects may have felt that they should not trust their immediate response to each performance, but they apparently had no consistent criteria for composer-specific qualities of expression.

EXPERIMENT 2

Experiment 1 had used a somewhat heterogeneous set of materials. Experiment 2 used a different, more controlled set of pieces, all Minuet movements from piano sonatas by the same four composers, selected by the author. Each movement had three sections: Minuet, Trio (or second Minuet), and repeat of the Minuet. Each section had a two-part structure, each part typically comprising 8 bars. In several ways, this choice of materials provided a rather extreme test of subjects' ability to appreciate composers' pulses. First, the basic meter was 3/4, so a 3-pulse was used for at least one level of the pulse hierarchy (the higher level, because of the moderately slow tempo of Minuets). Three-pulses have one degree of freedom less than 4-pulses and therefore are less effective in differentiating individual composers (Clynes, 1987). Second, the Minuet is a traditional dance form and thus not only imposes some constraints on the expression of a composer's individual characteristics, but also may require an additional "Minuet pulse" that was not implemented in the present materials (Clynes, personal communication). Third, the pieces were all similar in structure, tempo, duration, key signature, and sound quality. Finally, judging from Clynes' publications, the microstructure of the 3-pulses may not be as confidently established as that of the 4-pulses.

Experiment 2 was conducted in three versions, referred to as 2a, 2b, and 2c. Experiment 2a took advantage of the possibility that the Trio (or second Minuet), whose character was generally quite different from that of the (first) Minuet, could be considered as a separate composition. By conducting separate tests for Minuet and Trio, Experiment 2a examined whether the same pulse preferences would be obtained for these different parts and thus tested the hypothesis (prediction 4 in the introduction) that a composer's pulse should apply to all of his music. (For some qualifications of that hypothesis, see below.) In addition, the Minuet test was repeated to assess the reliability of subjects' judgments and possible effects of test sequence. The instructions were the same as in Experiment 1a. To counter possible objections against the separation of Minuet and Trio from a musical (rather than methodological) perspective (Clynes, personal communication), Experiment 2b replicated Experiment 2a using the integral Sonata movements as well as (in part) improved sound reproduction and modified instructions. Experiment 2c, finally, was a replication of Experiment 2b using amended versions of two pieces, further improvements in sound reproduction, and the instructions of Experiment 1b.

Methods

Subjects

Experiment 2a. Twelve subjects (4 women, 8 men) participated in this experiment. Eight of them (including the author), all with musical experience, were also subjects in Experiment 1a (which, chronologically, followed Experiment 2a). The other four subjects were two proficient amateur musicians (French horn, trumpet) and two relatively inexperienced individuals, all graduate students or young researchers. In this subject group, years of musical education ranged from 0 to 25, active playing hours from 0 to 11, and passive listening hours from 1 to 20. In addition, four professional pianists were tested. Three of them (one woman, two men) were young performers and teachers residing in New Haven; the fourth was an experienced, middle-aged, female piano instructor at a local music school.

Experiment 2b. Five of the most reliable amateur subjects participated here, all of whom had been subjects in Experiment 2a.

Experiment 2c. Eleven of the subjects of Experiment 1b, plus three new subjects from Trinity College, participated here.

Materials

The four pieces selected by the author were Minuet movements from piano sonatas by Beethoven, Haydn, Mozart, and Schubert. Thus they were all of similar form, duration, and time signature (3/4), and they even had similar key signatures, although this was an irrelevant fact. Specifically, they were:

Beethoven: Menuetto (Moderato e grazioso) from the Sonata in E-flat major, op. 31, No. 3 (with Trio in the same key and a Coda after the repeat of the Minuet);

Haydn: Menuetto (no tempo indication) from the Sonata in E-flat major, HV XVI/28 (with Trio in e-flat minor);

Mozart: Menuetto (Allegretto) in B-flat major from the Sonata in E-flat major, K. 282 (with Menuetto II in E-flat major);

Schubert: Menuetto (Allegretto) from the Sonata in E-flat major, op. 122, D. 568 (with Trio in A-flat major).

The initial bars of the Minuets and Trios (or second Minuet in the Mozart) are shown in Figure 2.

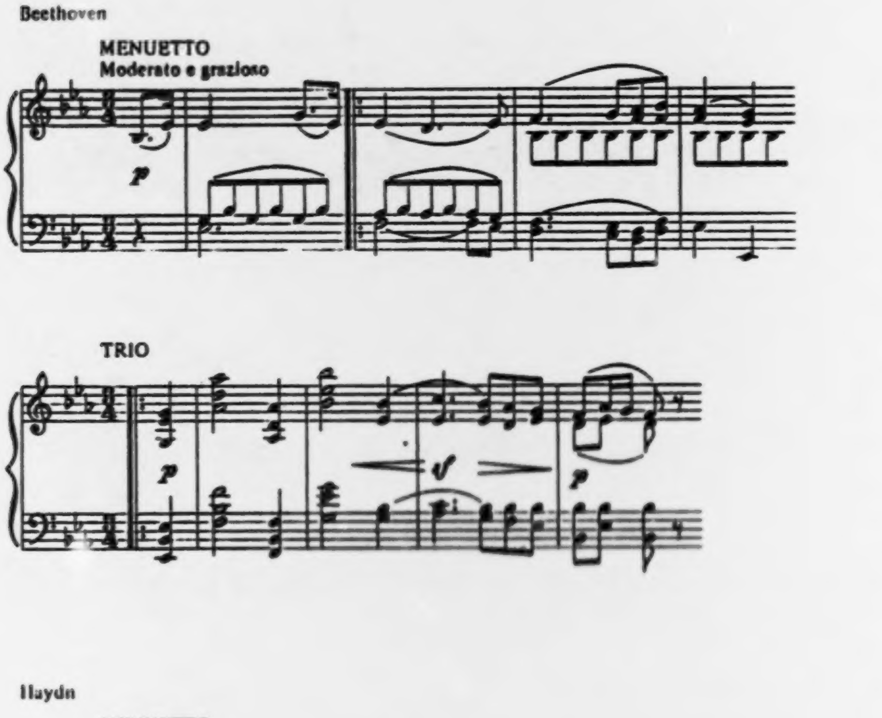
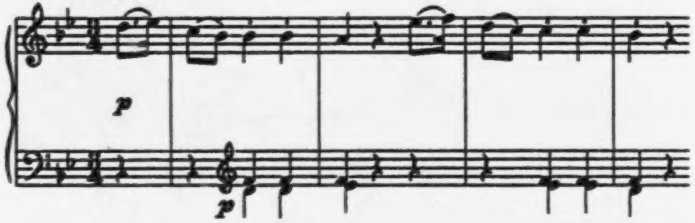

Beethoven
MENUETTO
Moderato e grazioso
p
TRIO
p
sf
p
Haydn
MENUETTO
f
TRIO
p

Figure 2. Initial bars of both parts of the pieces used in Experiment 2.

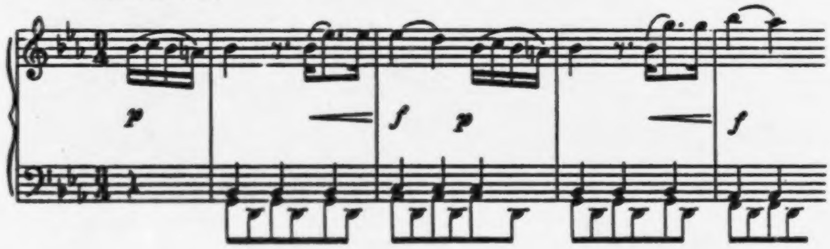
Mozart

MENUETTO I

Allegretto



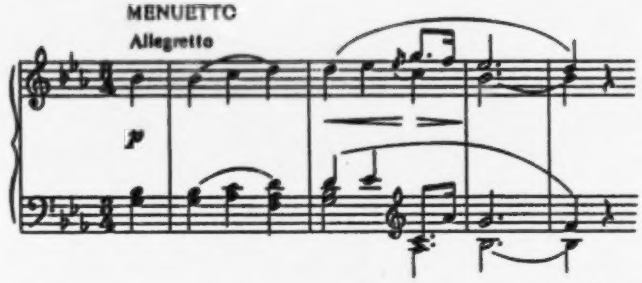
MENUETTO II



Schubert

MENUETTO

Allegretto



TRIO

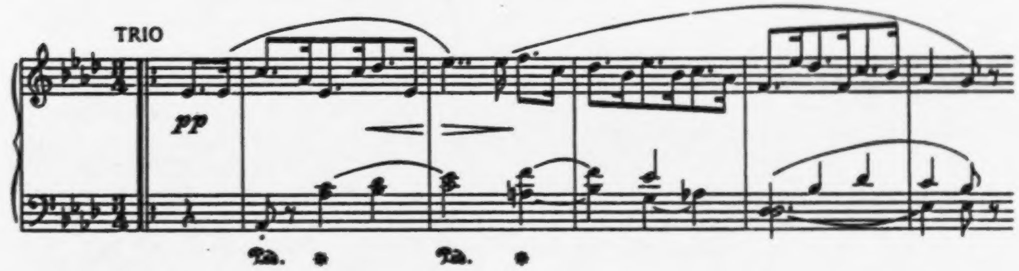


Figure 2 continued.

All pieces were generated at the author's request by Clynes in Sydney with Roland MKS-20 piano sound, without any repeats. As in Experiment 1, each piece was recorded (with Dolby B) in a neutral version and in four pulsed versions. The pulses used and their attenuation factors are shown in Table 5. It should be noted that although a 3-

pulse (see Table 1) was employed at the higher level (comprising the three quarter-notes within a bar), a 4-pulse was used at the lower level (sixteenth-notes). (For the treatment of triplets, see Clynes, 1987, p. 214.) Thus the lower-level pulses were largely the same as in Experiment 1, though there were far fewer notes at that level in the present Minuets.

TABLE 5. Hierarchical pulse patterns (see Table 1) and attenuation factors for the timing pulses used in synthesizing the pieces of Experiment 2 (Clynes, p.c.).

Composer	Lower level	Higher level
Beethoven	Four-pulse (1.0)	Three-pulse (0.5)
Haydn	Four-pulse (1.0)	Three-pulse (1.0)
Mozart	Four-pulse (1.0)	Three-pulse (0.75)
Schubert	Four-pulse (1.1)	Three-pulse (1.1)

Procedure

Experiment 2a. The experimental tape contained three tests, each structured like the single test in Experiment 1. The first test contained only the Minuets, the second the Trios (or second Minuet), and the third the Minuets again (identical except in the case of the Beethoven, where the Coda was included). Each test had a different order of composers and a different order of versions for each piece, following a different Latin square design in each test. As in Experiment 1, the neutral version always preceded the four pulsed versions. Instructions and answer sheets were the same as in Experiment 1a, with only minor changes reflecting the different materials. The procedures were also the same. The nonprofessional subjects were tested at Haskins and Yale using a portable stereo cassette recorder without Dolby B. Three professional pianists were tested in a studio or in their home using the same equipment; the fourth used her home stereo system. The repeat of the Minuet test was omitted for the professionals in view of their busy schedule.

Experiment 2b. Here a single test was employed, using a new Latin square design. Each presentation of each piece was to consist of Minuet, Trio (or second Minuet), and repeat of the Minuet (plus Coda in the Beethoven). The Beethoven and Haydn were available in this integral form on the master tape. For the Mozart and Schubert, however, the master tape omitted the Minuet repeat. In a first version of the test tape, the pieces were played in this form, that is, without the Minuet repeat in the Mozart and Schubert, using a portable cassette recorder without Dolby B and the same instructions as in Experiment 2a. Subsequently, because precisely the Mozart and Schubert pieces did not yield the predicted results, a new test tape was created in which the Minuet repeats of the Mozart and Schubert were dubbed in from the master tape. That second version was presented to the same subjects a few months later using a different portable stereo cassette recorder with Dolby B and attached mini-loudspeakers. (One subject listened at home using his stereo equipment, with Dolby B.) The instructions were modified somewhat in the direction of those employed in Experiment 1b by omitting the term "lively" and emphasizing the term "idiomatic" (defined explicitly as "expressive in a way appropriate for that composer").

Experiment 2c. A new test tape was created with the same test sequence as in Experiment 2b, but with newly recorded versions of the Mozart and Schubert pieces furnished by Clynes, which avoided some amplitude distortions that had occurred in the earlier versions due to a nonlinear response of the sound module at high intensities (Clynes, personal communication). The tapes were played back on equipment at Trinity

College with Dolby B. The instructions were those of Experiment 1b, which emphasized the criterion of "composer-appropriateness."

Results and Discussion

Experiment 2a

The overall repeated-measures ANOVA in this experiment included not only composer and pulse as factors but also test (Minuet, Trio, Minuet). As in Experiment 1a, there was a highly significant main effect of pulse, $F(3,33) = 13.45$, $p < .0001$, indicating that some pulses were generally preferred over others. In addition, there was a highly significant composer by pulse interaction, $F(9,99) = 5.64$, $p < .0001$, which shows that pulse preferences changed reliably with composer, as in Experiment 1a. Thus, prediction 1 was again upheld, even for pulse patterns that were less differentiated and operated on similar time scales. There was also a marginally significant main effect of composer, $F(3,33) = 3.15$, $p = .0377$, and a significant composer by test interaction, $F(6,66) = 6.75$, $p < .0001$; both are of little interest. Importantly, however, the triple interaction was highly significant, $F(18,198) = 3.82$, $p < .0001$, suggesting that pulse preferences changed not only with composer, but also between Minuets and Trios.

Since this last interaction may also have reflected, in part, a change in judgments between the two presentations of the Minuet, a separate analysis was conducted on the Minuet data only. Significant effects included the main effects of both composer and pulse, as well as the crucial composer by pulse interaction, $F(9,99) = 5.26$, $p < .0001$. Test had no main effect, interacted only weakly with composer, and engaged in a triple interaction that fell just short of significance, $F(9,99) = 1.96$, $p = .0515$. It may be concluded, therefore, that the pulse preference pattern did not change substantially between the two presentations of the Minuets. Since the two Minuet tests followed different Latin square designs, this result also means that there were no obvious artifacts of test order. These conclusions were further supported by a significant correlation ($r = 0.74$, $p < .01$) between the average ratings for the 16 pulsed versions in the two Minuet tests, which gives some indication of the reliability of the judgments of the subject group as a whole. Considering that there was only a single judgment per stimulus and subject, the reliability is quite satisfactory.

A separate analysis of the Trio data revealed, besides a pulse main effect, a significant composer by pulse interaction, $F(9,99) = 5.36$, $p < .0001$. This interaction thus held for both Minuet and Trio separately and combined, but its precise pattern was different for the two. This was confirmed by low and nonsignificant correlations between the average Trio ratings and the first and second Minuet ratings, respectively ($r = 0.18$ and 0.32).

The response patterns may be examined in Table 6. It is evident that prediction 2, that the correct pulses would receive the highest ratings, was not well supported by the present results. Only the Beethoven pulse "worked", both in the Beethoven Minuet and in the Trio. (Paradoxically, it was precisely the Beethoven pulse that caused problems in Experiment 1; note the differences in meter and tempo, however.) In none of the other pieces did the correct pulse come out first or even receive very positive ratings. Some clear preferences for *incorrect* pulses emerged in the Haydn Trio (Mozart pulse) and the Mozart Trio (Haydn pulse). Some striking dislikes must also be noted, especially for the Mozart pulse in the Beethoven Minuet and in the Schubert Trio, and for the Schubert pulse in the Haydn and Mozart Minuets. At least, no strong dislikes occurred for correct pulses anywhere.

TABLE 6. Average ratings of the computer performances in Experiment 2a (12 nonprofessional subjects).

Composer	Pulse			
	Beethoven	**Haydn**	**Mozart**	**Schubert**
		MINUET		
Beethoven	*1.33*	0.75	-3.25	-1.42
Haydn	0.58	*-0.08*	0.25	-2.33
Mozart	0.58	0.17	*0.17*	-1.33
Schubert	1.50	1.42	1.08	*0.75*
		TRIO		
Beethoven	*2.33*	0.33	0.92	0.08
Haydn	-1.00	*0.42*	1.83	-0.83
Mozart	0.17	1.42	*-0.50*	-1.08
Schubert	1.17	1.17	-1.67	*-1.08*
		MINUET (REPEAT)		
Beethoven	*1.33*	-0.42	-1.75	-0.42
Haydn	1.25	*0.58*	0.33	-1.33
Mozart	-0.83	0.33	*0.08*	-1.83
Schubert	0.75	1.42	0.83	*-0.42*

The patterns described above were all significant when tested for each composer separately: The pulse main effects were significant in each case, and the pulse by test interaction was significant for all pieces except the Mozart. Thus there were reliable dependencies of pulse preferences on composers *and* pieces (Minuet vs. Trio), which implies that the pulses do not apply equally to all parts of a musical composition.

Even though there was little support for the specific predictions overall, an overall PAI was nevertheless computed for each subject, and its relation to the three measures of musical experience was examined, just as in Experiment 1. Half the subjects had positive PAIs, half had negative ones. While the correlations with years of musical education and hours of passive listening were negligible, the correlation with active playing hours per week reached significance ($r = 0.61$, p .05). This correlation reinforced a weak trend in the same direction observed in Experiment 1a, suggesting that listeners with more active musical experience are more appreciative of the correct pulses. This raised the question of whether professional musicians might give judgments that are more in conformity with the predictions.

The average judgments of the four professional pianists are shown in Table 7. Because of the small number of subjects, only a qualitative comparison can be made with the data in Table 6. Clearly, there are some similarities and some differences. In the Beethoven, the professionals, too, showed a consistent preference for the Beethoven pulse. A striking difference from the earlier data is their liking of the Schubert pulse in the Beethoven Trio, to which the nonprofessional subjects had been indifferent. Their opposite reactions to the Schubert pulse in the Beethoven Minuet and Trio are interesting. In contrast to the nonprofessional subjects' results, the Haydn pulse came out best in the Haydn Minuet, but not impressively so, as it was not much preferred over the neutral pulse. (Also, only one subject actually gave the Haydn pulse the highest rating; each of the other three subjects preferred a different pulse, and one of them ranked the Haydn pulse last.) In the Haydn Trio, the Haydn pulse ranked second to the Schubert pulse, but neither was preferred over the neutral version. The professionals disliked the Mozart pulse in the Haydn Trio, in stark contrast to the nonprofessional subjects. No clear preferences emerged in the Mozart pieces, but in both the Mozart pulse ranked last. In the Schubert Minuet, the Schubert pulse, which had ranked last with the nonprofessional subjects, came out first. (However, only one of the four professionals actually preferred the Schubert pulse, for two it was tied with other pulses for first place, and one ranked it last.) In the Schubert Trio, the Beethoven pulse was clearly preferred.

TABLE 7. Average ratings of the computer performances in Experiment 2a (4 professional pianists).

Composer	Pulse			
	Beethoven	Haydn	Mozart	Schubert
			MINUET	
Beethoven	*0.75*	-1.25	-4.00	-3.50
Haydn	0.25	*0.50*	-2.00	-2.00
Mozart	-0.50	-1.25	*-1.75*	-0.75
Schubert	0.00	-0.25	-2.00	*0.75*
			TRIO	
Beethoven	*3.00*	1.25	0.00	.50
Haydn	-1.25	*-0.25*	-1.75	0.00
Mozart	0.00	0.00	*-1.50*	-1.00
Schubert	1.25	-1.75	-3.00	*-1.25*

In summary, the four professional musicians' results yielded a somewhat more positive response to the Haydn and Schubert pulses in the Minuets, but not in the Trios. The Mozart pulse remained unsuccessful throughout. There were also large individual differences among the professionals. In terms of the PAI, the measure of individual conformity to prediction 2, two of the four professionals would have ranked among the top four subjects, had they been included with the other subjects; the other two, however, would have ranked at the low end. Therefore, inclusion of their data would not have increased the correlation between the PAI and musical experience.[9]

Several additional correlations were computed for the 12 nonprofessional subjects. Their average ratings (i.e., their overall tendency to give positive ratings, or to dislike the neutral version) correlated positively with years of music instruction but negatively with active playing hours ($r = -0.51$, $p < .10$) and passive listening hours. Though none of these correlations reached conventional levels of significance, they support a tendency observed in Experiment 1a for musically more active subjects to like pulsed versions less. This tendency is further supported by the results of the professional musicians (see Table 7), who tended to give even more negative ratings overall. Perhaps, these subjects expected more "interpretation" from the pulsed versions than they actually provided. The subjects' familiarity ratings were also examined. A somewhat wider range than in Experiment 1a was found, though the pieces were again relatively unfamiliar to most subjects. There was no correlation between familiarity and the average PAI (averaged over Minuet and Trio). Finally, for those eight subjects who participated in both Experiments 1a and 2a, the correlation of PAIs across the two experiments was computed and found to be significant ($r = 0.74$, $p < .02$). Thus there seemed to be reliable individual differences in the extent to which subjects gave judgments in agreement with prediction 2—that is, in the degree to which they agreed with Clynes.

Experiment 2b

One possible concern with the results of Experiment 2a was that the Trios might be less characteristic of a composer's style, and that their separation from the Minuet may have disrupted the continuity of the musical composition and impeded listeners' appreciation of the pulses. Of course, the Minuet results in the first third of the test cannot be explained away in this fashion. However, it is conceivable that judgments of the Trio would change when it follows immediately upon the Minuet; and it could also be that subjects modify their judgments of the Minuet after hearing the Trio. It is also possible that the absolute duration of the music plays a role in stabilizing the pulse. Therefore, Experiment 2b presented the integral Sonata movements. In the first version of the test, the Minuet repeats of the Mozart and Schubert were missing. In the second version, all pieces were complete, and there were modified instructions and Dolby B playback.

The data from both versions of the test (presented to the same group of subjects) were submitted to a repeated-measures ANOVA with the factors version, composer, and pulse. There were two significant effects: a main effect of pulse, $F(3,12) = 7.79$, $p = .0038$, due to a general preference for the Haydn pulse, followed by Beethoven, Mozart, and Schubert, and a composer by pulse interaction, $F(9,36) = 4.09$, $p = .0011$. None of the effects involving versions was significant; thus it may be concluded that addition of Minuet repeats in two of the pieces, changes in instruction, and Dolby B had little effect on subjects' judgments.

The results are shown in Table 8. They are not unlike the average results of the 12 subjects in Experiment 2a (cf. Table 6), with one striking exception: The Haydn pulse suddenly emerged as a clear winner in the Haydn piece. The magnitude of the effect seems almost miraculous; since it was present in both versions of Experiment 2b, it can only be attributed to the integrity of the Sonata movement. For none of the other three pieces, however, did the integrity of the composition have a similarly enhancing effect on the appreciation of the correct pulse. In fact, in the Beethoven piece, the Beethoven pulse was now liked less than the Haydn pulse, and the Mozart and Schubert pulses remained quite ineffective for their respective composers.

Even though the statistical analysis did not reveal any significant differences between the two versions of Experiment 2b, this could have been because of the small number of subjects. Therefore, the average ratings for the two versions were also compared by eye. Compared to version 1, the ratings for version 2 were lower overall; this difference actually approached significance. This was particularly true for the correct pulses in all four pieces; there was no indication whatsoever that any correct pulse was appreciated more in version 2 than in version 1. Thus the experiment was entirely negative with respect to possible effects of addition of the Minuet repeat in the Mozart and Schubert, instructions, and Dolby B playback.

TABLE 8. Average ratings of the computer performances in Experiment 2b.

	Pulse			
Composer	**Beethoven**	**Haydn**	**Mozart**	**Schubert**
		FIRST VERSION OF TEST		
Beethoven	*0.83*	1.00	-2.83	0.33
Haydn	-0.50	*3.00*	-0.17	-2.83
Mozart	0.67	0.33	*-0.17*	-2.67
Schubert	1.33	1.67	-0.83	*0.67*
		SECOND VERSION OF TEST		
Beethoven	*0.67*	1.00	-1.50	0.00
Haydn	0.67	*2.17*	-0.67	-2.33
Mozart	-0.17	-0.67	*-0.17*	-2.00
Schubert	0.67	0.17	-0.17	*-2.00*

Experiment 2c

The results of this further replication, with modified instructions and technically improved versions of the Mozart and Schubert pieces, are shown in Table 9. In the Beethoven piece, a preference for the Beethoven pulse emerged again, as in Experiment 2a. In the Haydn piece, the Beethoven pulse was preferred somewhat over the Haydn pulse, which also resembles the results for the Haydn Minuet in Experiment 2a (nonprofessional subjects, Table 6); the striking preference for the Haydn pulse obtained in Experiment 2b (Table 8) was not replicated, even though the integrality of the composition was preserved. In the Mozart piece, there was a marginal preference for the Mozart pulse, though it was not rated much above the neutral version. Relatively speaking, this constitutes a slight improvement over the previous results. Finally, in the Schubert piece the Schubert pulse received a positive rating but ranked behind the Haydn and Beethoven pulses. This may also be taken as a slight improvement, but it is certainly not impressive.

TABLE 9. Average ratings of the computer performances in Experiment 2c.

	Pulse			
Composer	**Beethoven**	**Haydn**	**Mozart**	**Schubert**
Beethoven	1.71	-0.50	-1.00	0.50
Haydn	1.57	1.21	0.50	0.50
Mozart	-1.00	0.29	0.43	-0.50
Schubert	1.50	1.57	0.43	1.36

As in Experiment 1b, which employed the same instructions with largely the same subjects, the results were less consistent than in previous runs of the test. In the ANOVA there was a significant main effect of composer, $F(3,39)$ 5.51, $p = 0.0030$, due to higher ratings for Schubert and Haydn than for Beethoven and Mozart. There was no significant main effect of pulse, but the crucial composer by pulse interaction did reach significance, $F(9,117)$ 2.47, $p = 0.0130$. As in Experiment 2b, there were no significant correlations between average ratings, familiarity ratings, PAI scores, and the several indices of musical experience.

GENERAL DISCUSSION

The results of these very preliminary experiments establish quite clearly that listeners' relative preferences for different pulse patterns vary with the piece they are listening to. Whether or not this composition by pulse interaction is caused by listeners' appreciation of composer-specific characteristics, it is an important finding in itself. It indicates that the listeners judged the time-amplitude deviations in relation to the musical content of each piece (i.e., as the expressive variations they were intended to be), not just as physical deviations of varying magnitude. Their judgments presumably represent some measure of the degree to which the pulses fit the musical structure.

The pulses devised by Clynes were, of course, intended to provide an optimal fit for each composer. Experiment 1a indeed revealed largely the expected pattern of pulse preferences (with only the Beethoven pulse being not convincing), though the results of Experiment 1b were less encouraging. Nevertheless, these data may be taken to provide some support for Clynes' choice of time-amplitude patterns for quadruple meter, at least for Mozart and Schubert. Curiously, the results of Experiment 2 were complementary to those of Experiment 1 in that they provided clear support for the Beethoven pulse and occasional support for the Haydn pulse, but little evidence in favor of the Mozart and Schubert pulses. Moreover, in Experiment 2a listeners' judgments differed for the Minuet and Trio parts.

The majority of the present subjects were moderately competent judges of the general quality of a musical performance. They may be considered representative of that valuable subpopulation of concert (or radio) audiences whose members actually attend to the music and to the manner in which it is performed. Moreover, their judgments were reasonably consistent; clearly, they were not just guessing. Possible weaknesses in the methodology (memory requirements, less than perfect sound reproduction) cannot be held responsible for consistent response patterns, only for noise in the data. Although an inappropriate pulse (i.e., that of a composer other than the composer of the piece) may occasionally have a positive effect, it is not clear why the appropriate pulse microstructure should ever be *less* appealing than an inappropriate or neutral one to the moderately sophisticated listener. Such reversals of judgment may be caused by *general* pulse preferences or aversions, but the negative findings in the present study cannot be explained on these grounds. Therefore, these findings indicate problems with the pulses themselves and their implementation, or possibly with the musical pieces chosen.

As to the pulses themselves, Clynes (1983, 1987) has stated that they are subject to continuing refinement and improvement (though presumably by artists like, not by the consensus of ordinary music lovers). The 3-pulses especially were called "tentative" by Clynes (1983), though some have been revised by him in the meantime. These patterns are still not optimal, and perhaps they have some idiosyncratic features that are unacceptable to other listeners. Each of the two independent parameters of the pulse, duration and amplitude, has a certain depth of modulation; that is, each pulse is a member of a whole family of similar patterns varying in the magnitude of the prescribed deviations. Even if the basic pattern is right, the modulation depth may be either too large or too small, with the resulting pulse either sounding exaggerated or being ineffective. It may be noted, in that connection, that Clynes developed the pulses using sine wave synthesis, whereas most of the present pieces were performed with a synthetic piano sound. The different sound quality, with its different overtone structure and spectral balance, may require an adjustment in the pulse modulation depth. The pulse may also need to adjust somewhat to different tempos, overall loudnesses, and the presence of multiple voices. Clynes' "attenuation factors" attempted to achieve these goals, but perhaps not with full success.

Another relevant issue concerns the maintenance of the pulse throughout a piece. As pointed out in the introduction, this follows from Clynes' earlier ideas on "time-form printing" in the central nervous system and is an important part of the theory. Clynes (1986) links it to psychobiological clocks and repetitive motor activities such as walking; he claims that listeners expect repetition of the pattern and would consider substitution of a different pattern disturbing. Essentially, the physical pulse, as instantiated in computer performances, is modelled after the *inner* (mental, subjective) pulse, which (according to Becking's, 1928, and Clynes' introspections) does possess the postulated constancy. However, Clynes (1987) also notes that actual performances are likely to contain "noise" in the form of random and planned deviations from the pulse, as one should expect from a human performer. Also, Becking (1928) commented that a composer's characteristic dynamics are not expressed equally throughout a composition, so the modulation depth of a pulse may have to vary in the course of a piece. In addition, of course, there are many other expressive devices (such as crescendi and diminuendi, special accents, ritardandi and accelerandi, phrase-final lengthening, etc.) that would be laid on top of a pulse in an actual performance. Few of these devices were used in the present computer performances, so as not to introduce too many complexities. However, the result may have been that the pulses were too obvious and exposed to be wholly pleasing. The precise, naked repetition of the pulse identifies the performance as that of a machine, even though the pulse itself is intended to impart "living qualities." While expert listeners (such as Clynes himself) may be trained to evaluate such bare pulses in terms of the musical thought they evoke, other listeners may appreciate pulses, if at all, only as the background carrier for other, structurally determined modulations of the musical flow. All this goes to say that the present materials were deliberately simple performances for laboratory use, and that perhaps a sweeping success of the pulses with untrained listeners should not have been expected, particularly in Experiment 2, where the Minuet dance characteristics were deliberately absent.

Finally, we turn to the issue of the generality of the pulses across all works of a composer. Although Clynes has likened composers' pulses to stable individual characteristics such as handwriting or voice quality (Clynes, 1986) and has stated that they do not seem to change with a composers' age (Clynes, 1987), in personal communications he has expressed reservations about some of the present pieces on the grounds that they represent compositions from the composer's youth (e.g., Beethoven's op. 10, No. 2, used in Experiment 1). He has also pointed out to this author that, in his opinion, the Trio sections of Minuet movements are often not characteristic of their composers' style, and that this may explain some of the poor results in Experiment 2a. However, while random results would be consistent with this suggestion, systematic preferences for inappropriate pulses are more difficult to explain. Moreover, the suggestion that a pulse applies only to music that is characteristic of its composer threatens to make the theory circular, since there is no objective measure of "characteristicness." An unlimited number of pieces could be exempted on these grounds. Similarly, it is not clear where to draw the line between early and mature compositions. (Mozart, for example, had written almost 300 other works before the piano sonatas from which the present excerpts were taken, and he surely had developed his personal style by that time.) It would be quite reasonable to amend Clynes' theory by allowing composers' pulses to become more pronounced and clearly defined with age; thus listeners may have greater difficulty appreciating the "correct" pulse in earlier compositions. It seems, however, that even such an amended theory would not allow for preferences of *other* composers' pulses over the correct one, as occurred in several instances.

The generality of a composer's pulse over his whole oeuvre is clearly implied by Clynes' theory as presently stated. Obviously, a much larger set of compositions will be needed to test this part of the theory thoroughly. However, it is clear that the idea of a single characteristic pulse for a composer is an *abstraction* derived from intimate familiarity with his total oeuvre. This ideal pulse can perhaps be appreciated by artists who are able to experience any given piece in the context of stored knowledge about the composer's total output. Knowing that a piece is by Beethoven, say, they would find the Beethoven pulse appropriate because they *recognize* it as "Beethovenian." Individual compositions of a composer, however, may require an expressive microstructure that deviates more or less from the ideal composer's pulse, in accordance with the specific structural properties of the piece. (An analogy to the mean and variance of a statistical distribution might be appropriate.) Depending on the composer's expressive range and stylistic consistency, that piece-specific pulse may be a variant of the ideal pulse, or it even may be in conflict with it. Moreover, it may vary in the course of a piece as the musical structure unfolds.[10] The subjects in the present experiments may have judged the performances in such a piece-specific frame of reference, despite the revised instructions in Experiments 1b and 2c. If so, their ratings may indeed be a measure of the "typicality" of a piece among all compositions of a composer, assuming that Clynes' pulse patterns are close to the true ideal. The negative results in Experiment 2, for example, could be interpreted as suggesting that the Mozart and Schubert Minuet movements were not typical of their respective composers, that the Mozart Trio could be mistaken for Haydn, the Haydn Trio for Mozart, and so on. Some independent perceptual or musicological criterion will be needed to judge whether these suggestions are tenable.

ACKNOWLEDGMENT

This research would not have been possible without the collaboration, advice, and criticism of Manfred Clynes, to whom I am deeply indebted. His extensive suggestions also helped improve the present manuscript. Helpful comments on earlier versions of the manuscript were also obtained from Robert Crowder, Margaret Dunn, Carol Krumhansl, Mary Louise Serafine, and William Thompson. I am especially grateful to Eve Perugini, who ran Experiments 1b and 2c as part of her honor's thesis at Trinity College, Hartford, Connecticut, and to my colleagues at Haskins and Yale and to the professional musicians who donated their time to serve as subjects. Some financial support for this project came from NIH BRS Grant RR-05596 to Haskins Laboratories. A short version of the paper was presented at the 113th meeting of the Acoustical Society of America in Indianapolis in May 1987.

REFERENCES

Becking, G. (1928). *Der musikalische Rhythmus als Erkenntnisquelle*. Augsburg: Benno Filser.

Bengtsson, I., & Gabrielsson, A. (1980). Methods for analyzing performance of musical rhythm. *Scandinavian Journal of Psychology, 21*, 257-268.

Clynes, M. (1969). Toward a theory of man. Precision of essentic form in living communication. In K. N. Leibovic & J. C. Eccles (Eds.), *Information processing in the nervous system* (pp. 177-206). New York: Springer-Verlag.

Clynes, M. (1977). *Sentics. The touch of emotions.* Garden City, NY: Anchor Press/Doubleday.

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance* (pp. 76-181). Publication issued by the Royal Swedish Academy of Music No. 39 (Stockholm, Sweden).

Clynes, M. (1985). Secrets of life in music. In *Analytica. Studies in the description and analysis of music in honor of Ingmar Bengtsson* (pp. 3-15). Publication issued by the Royal Swedish Academy of Music No. 47 (Stockholm, Sweden).

Clynes, M. (1986). Generative principles of musical thought: Integration of microstructure with structure. *Communication and Cognition AI, 3,* 185-223.

Clynes, M. (1987). What can a musician learn about music performance from newly discovered microstructure principles (PM and PAS)? In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 201-233). Publication issued by the Royal Swedish Academy of Music No. 55 (Stockholm, Sweden).

Clynes, M., & Walker, J. (1982). Neurobiologic functions of rhythm, time, and pulse in music. In M. Clynes (Ed.), *Music, mind, and brain: The neuropsychology of music.* New York: Plenum Press.

Gabrielsson, A., Bengtsson, I., & Gabrielsson, B. (1983). Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology, 24,* 193-213.

Hartmann, A. (1932). Untersuchungen über mètrisches Verhalten in musikalischen Interpretationsvarianten. *Archiv für die gesamte Psychologie, 84,* 103-192.

Seashore, C. E. (1938/1967). *Psychology of music.* New York: McGraw-Hill (original), Dover (reprint).

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology, 13,* 326-376.

Shaffer, L. H., Clarke, E. F., and Todd, N. P. (1985). Metre and rhythm in piano playing. *Cognition, 20,* 61-77.

Sievers, E. (1924). *Ziele und Wege der Schallanalyse.* Heidelberg: Carl Winter.

Sundberg, J., Frydén, L., & Askenfelt, A. (1983). What tells you the player is musical? In J. Sundberg (Ed.), *Studies of music performance* (pp. 61-75). Publication issued by the Royal Swedish Academy of Music No. 39 (Stockholm, Sweden).

Sundberg, J., & Frydén, L. (1985). Teaching a computer to play melodies musically. In *Analytica. Studies in the description and analysis of music in honour of Ingmar Bengtsson* (pp. 67-76). Publications issued by the Royal Swedish Academy of Music No. 47 (Stockholm, Sweden).

Thompson, W. F. (in press). Musical performance and the composer: A consideration of Clynes' theory of pulse for performances of Mozart and Beethoven. *Music Perception.*

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3,* 33-58.

FOOTNOTES

[*] *Music Perception, 6,* 243-274 (1989).

[1] The term "duration" is used by Clynes, though it is meant to refer not to the time from the onset of a note to its offset (which depends on the degree of legato or staccato) but to the time from the onset of one note to the onset of the next note, or the *onset-to-onset interval* (OOI).

[2] Part of a "Computerized System for Imparting an Expressive Microstructure to Succession of Notes in a Musical Score" that was awarded U.S. Patent No. 4,704,682 (November 3, 1987).

[3] Shorter notes within a pulse beat are treated according to a pulse at that level (Clynes, personal communication).

[4] Nevertheless, the fact that the present materials (with one exception) had not been tried out previously with composers' pulses made these experiments a test of the generality of the pulses.

[5] Although the author had made up the stimulus tapes and thus was the only subject to have heard the materials before, he had no recollection of the test order at the time of testing and had not yet learned to identify the individual pulse patterns.

[6] Because of these attenuation factors, the higher-level pulses were not strictly the same across the different compositions. This methodological complication had to be ignored for purposes of analysis. It might also be noted that the basic pulse patterns for different composers (Table 1) are not equal with respect to average modulation depth.

[7] The Mozart-pulse version of the Beethoven, however, was recorded in regular piano sound; this did not seem to affect listeners' judgments. The other versions were also used by Thompson (submitted, Exp. 3) and can be heard on the record accompanying Clynes (1987).

[8] This was also true when years of musical education were computed for the dominant instrument only or for piano only. To illustrate some of the individual variability: The subject with the highest PAI (3.00) is an active amateur cellist; the person with the second-highest PAI (2.52) never had any music education; the person with the lowest PAI (-1.33) had the longest music education of all, though she no longer plays her instruments (piano and violin).

[9] It was not attempted to estimate numerically the musical experience of the professionals. Clearly, it was far above that of the other subjects; therefore, computation of correlations across all subjects was not advisable.

[10] It would exceed the scope of the present study to investigate this hypothesis further, in view of the length and complexity of the musical materials. It was observed informally by the author and other listeners that the positive or negative impressions evoked by a pulse did vary in the course of a piece. A more detailed investigation of the interaction of structural properties and pulse effectiveness would require tracking listeners' judgments continuously or presenting shorter, structurally homogeneous musical excerpts for evaluation.

Appendix

SR #	Report Date	DTIC #	ERIC #
SR-21/22	January-June 1970	AD 719382	ED 044-679
SR-23	July-September 1970	AD 723586	ED 052-654
SR-24	October-December 1970	AD 727616	ED 052-653
SR-25/26	January-June 1971	AD 730013	ED 056-560
SR-27	July-September 1971	AD 749339	ED 071-533
SR-28	October-December 1971	AD 742140	ED 061-837
SR-29/30	January-June 1972	AD 750001	ED 071-484
SR-31/32	July-December 1972	AD 757954	ED 077-285
SR-33	January-March 1973	AD 762373	ED 081-263
SR-34	April-June 1973	AD 766178	ED 081-295
SR-35/36	July-December 1973	AD 774799	ED 094-444
SR-37/38	January-June 1974	AD 783548	ED 094-445
SR-39/40	July-December 1974	AD A007342	ED 102-633
SR-41	January-March 1975	AD A013325	ED 109-722
SR-42/43	April-September 1975	AD A018369	ED 117-770
SR-44	October-December 1975	AD A023059	ED 119-273
SR-45/46	January-June 1976	AD A026196	ED 123-678
SR-47	July-September 1976	AD A031789	ED 128-870
SR-48	October-December 1976	AD A036735	ED 135-028
SR-49	January-March 1977	AD A041460	ED 141-864
SR-50	April-June 1977	AD A044820	ED 144-138
SR-51/52	July-December 1977	AD A049215	ED 147-892
SR-53	January-March 1978	AD A0[illegible]5853	ED 155-760
SR-54	April-June 1978	AD A067070	ED 161-096
SR-55/56	July-December 1978	AD A065575	ED 166-757
SR-57	January-March 1979	AD A083179	ED 170-823
SR-58	April-June 1979	AD A077663	ED 178-967
SR-59/60	July-December 1979	AD A082034	ED 181-525
SR-61	January-March 1980	AD A085320	ED 185-636
SR-62	April-June 1980	AD A095062	ED 196-099
SR-63/64	July-December 1980	AD A095860	ED 197-416
SR-65	January-March 1981	AD A099958	ED 201-022
SR-66	April-June 1981	AD A105090	ED 206-038
SR-67/68	July-December 1981	AD A111385	ED 212-010
SR-69	January-March 1982	AD A120819	ED 214-226
SR-70	April-June 1982	AD A119426	ED 219-834
SR-71/72	July-December 1982	AD A124596	ED 225-212
SR-73	January-March 1983	AD A129713	ED 229-816
SR-74/75	April-September 1983	AD A136416	ED 236-753
SR-76	October-December 1983	AD A140176	ED 241-973
SR-77/78	January-June 1984	AD A145585	ED 247-626
SR-79/80	July-December 1984	AD A151035	ED 252-90
SR-81	January-March 1985	AD A156294	ED 257-159
SR-82/83	April-September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022
SR-86/87	April-September 1986	AD A176816	ED 278-066
SR-88	October-December 1986	PB 88-244256	ED 282-278

SR-89/90	January-June 1987	PB 88-244314	ED 285-228
SR-91	July-September 1987	AD A192081	**
SR-92	October-December 1987	PB 88-246798	**
SR-93/94	January-June 1988	PB 89-108765	**
SR-95/96	July-December 1988	PB 89-155329/AS	
SR-97/98	January-June 1989		

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corporation (CMC)
3900 Wheeler Avenue
Alexandria, VA 22304-5110

In addition, Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, CA 92122

**Accession number not yet assigned

Haskins Laboratories Status Report on
Speech Research

SR-97/98
JANUARY-JUNE 1989

Contents

- **Gestural Structures and Phonological Patterns**
  Catherine P. Browman and Louis Goldstein 1
- **The Cricothyroid Muscle in Voicing Control**
  Anders Löfqvist, Thomas Baer, Nancy S. McGarr, and Robin Seider Story 25
- **Growth of Surface Waves with Application to the Mucosal Wave of the Vocal Folds**
  R. S. McGowan 41
- **Effects of Preceding Context on Discrimination of Voice Onset Times**
  Bruno H. Repp and Hwei-Bing Lin 51
- **Can Speech Perception be Influenced by Simultaneous Presentation of Print?**
  Ram Frost, Bruno H. Repp, and Leonard Katz 69
- **Detectability of Words and Nonwords in Two Kinds of Noise**
  Bruno H. Repp and Ram Frost 87
- **Discovering Phonetic Coherence in Acoustic Patterns**
  Catherine T. Best, Michael Studdert-Kennedy, Sharon Manuel, and Judith Rubin-Spitz 95
- **Reception of Language in Broca's Aphasia**
  Donald Shankweiler, Stephen Crain, Paul Gorrell, and Betty Tuller 121
- **Syntactic Comprehension in Young Poor Readers**
  Suzanne T. Smith, Paul Macaruso, Donald Shankweiler, and Stephen Crain 147
- **Phonological Deficiencies: Effective Predictors of Future Reading Problems**
  Virginia A. Mann and Patricia Ditunno 171
- **Consideration of the Principles of the International Phonetic Alphabet**
  Arthur S. Abramson 191
- **Language Specificity in Lexical Organization: Evidence from Deaf Signers' Lexical Organization of ASL and English**
  Vicki L. Hanson and Laurie B. Feldman 195
- **Anomalous Bimanual Coordination Among Dyslexic Boys**
  Marshall Gladstone, Catherine T. Best, and Richard J. Davidson 213
- **Expressive Microstructure in Music: A Preliminary Perceptual Assessment of Four Composers' "Pulses"**
  Bruno H. Repp 235

Appendix 261